International Journal of Human-Computer Interaction

Does augmented intelligence improve employee workload in decision-making situations? A systematic review

--Manuscript Draft--

Manuscript Number:	IJHC-D-23-00009
Full Title:	Does augmented intelligence improve employee workload in decision-making situations? A systematic review
Article Type:	Research Article
Section/Category:	Basic Science Section
Manuscript Region of Origin:	
Abstract:	According to the design approach of augmented intelligence, applications based on artificial intelligence (AI) are intended to complement humans in complex decision-making situations. Technological assistance should lead to a number of benefits for employees, such as less psychological overload in decision-making situations and better task performance. To the best of our knowledge, there has been no overview to date that summarizes whether and under what conditions the desired effects on employees are achieved. Therefore, this study identified and summarized articles from 2012 onwards that examined the effects of AI-based decision support systems on employees in controlled experimental studies. Through an exploratory search of the EBSCOhost database, 11,578 articles were found, of which 16 were ultimately identified as appropriate. Specific recommendations for action by researchers and practitioners were discerned from the results.

Does augmented intelligence improve employee workload in decision-making situations? A systematic review

Katharina Buschmeyer[a*], Julie Zenner[b], & Sarah Hatfield[a]

[a]*Faculty of Business, Augsburg University of Applied Sciences, Augsburg, Germany;* [b]*Faculty of Liberal Arts and Sciences, Augsburg University of Applied Sciences, Augsburg, Germany*

*Corresponding Author: Katharina Buschmeyer, Friedberger Straße 4, 86161 Augsburg, Germany; Email: katharina.buschmeyer@hs-augsburg.de

Katharina Buschmeyer is a junior researcher in the field of occupational psychology at the Augsburg University of Applied Sciences and Ph.D. student at the Ruhr University Bochum in Germany. Her research focuses on the use of augmented intelligence systems in the professional context to improve working conditions and the experience and behavior of employees. She holds a master's degree in business psychology (2019). ORCID: https://orcid.org/0000-0003-1625-7348; LinkedIn: https://www.linkedin.com/in/katharina-buschmeyer-55b325166.

Julie Zenner has a doctoral degree in the field of pedagogical psychology from the University of Siegen in Germany (2020). In her work, she focuses on the interaction between relationship experiences and career aspirations. Her current research interest is in the field of human-machine interaction at the Augsburg University of Applied Sciences. ORCID: https://orcid.org/0000-0001-8283-3097

Sarah Hatfield is professor of change management and human resources at the Augsburg University of Applied Sciences. She is the founder and head of the university's B.Sc. in Business Psychology and holds a diploma in work, organization and business psychology. Her research focuses on human-AI interaction, learning in virtual reality settings, and gamified learning. ORCID: https://orcid.org/0000-0003-1625-7348; https://www.linkedin.com/in/sarah-hatfield-2babaa181/

Does augmented intelligence improve employee workload in decision-making situations? A systematic review

According to the design approach of augmented intelligence, applications based on artificial intelligence (AI) are intended to complement humans in complex decision-making situations. Technological assistance should lead to a number of benefits for employees, such as less psychological overload in decision-making situations and better task performance. To the best of our knowledge, there has been no overview to date that summarizes whether and under what conditions the desired effects on employees are achieved. Therefore, this study identified and summarized articles from 2012 onwards that examined the effects of AI-based decision support systems on employees in controlled experimental studies. Through an exploratory search of the EBSCO*host* database, 11,578 articles were found, of which 16 were ultimately identified as appropriate. Specific recommendations for action by researchers and practitioners were discerned from the results.

Keywords: augmented intelligence; artificial intelligence; decision-making; decision support systems; employees; psychological workload; workload consequences

Introduction

Anyone dealing with the future of work sooner or later contemplates the relevance of artificial intelligence (AI). The computer science subfield has reached a new level of maturity in recent years through impressive improvements in sensors, data quality and quantity, memory, and computing capacity (Nicodeme, 2020). AI methods have enormous potential to increase profitability and maintain the competitiveness of companies by opening up efficiency gains and new business models (Benbya et al., 2020; Wamba-Taguimdje et al., 2020). Therefore, they are increasingly finding their way into businesses and our professional lives.

In the professional context, two concepts of employing AI-based solutions are often distinguished: (1) AI-based applications that support employees in completing their tasks, often referred to as *augmented intelligence*; and (2) AI-based applications that replace employees and are designed to function completely independently of human intervention, often referred to as *autonomous intelligence* (Hassani et al., 2020; Walch, 2020). In recent years, an increasing number of studies have focused on the former design approach (see, e.g., Jain et al., 2021; Jarrahi, 2018; Zheng et al., 2017), as do we in this study. This is because on the one hand, AI-based systems that keep humans in the loop seem to be more ethical and, therefore, desirable. On the other hand, in practice, fully automated AI-based applications are only suitable for a limited scope of use. As the complexity of the application domain increases, humans are becoming increasingly important for the interpretation and verification of AI-based system outputs (Hellebrandt et al., 2021; Zheng et al., 2017). In this sense, the design approach of augmented intelligence pursues a synergetic complementation of human and machine intelligence (Kirste, 2019), which often looks like this: AI-based applications analyze large amounts of data that are almost impossible for humans to grasp and provide the findings as a decision aid. The human being is then free to decide whether to use the system output as a basis for decision-making, that is, whether to follow the system's advice or not; thus, they still have sovereignty over the final decision and associated actions. In some cases, humans "feed" the system with their expertise and practical knowledge at the end of the decision-making process so that the system can be continuously improved (Hellebrandt et al., 2021).

This form of human-machine interaction is intended to create a win-win situation for companies and employees. Organizations should have a more efficient and less error-prone decision-making process. Employees are supposed to be cognitively relieved by AI-enabled support in decision-making and, thus, should experience less psychological overload when processing tasks, feel more confident in their decisions, and make better decisions. From a psychological perspective, better decision-making is particularly relevant; therefore, the question to be answered is whether the desired effects are actually achieved among employees. This human-centered view is increasingly being considered (see e.g., Geyer et al., 2022; Oppermann et al., 2019) as AI is progressively being viewed from a transdisciplinary perspective that integrates engineering, economics, and the social sciences (Y. Cai et al., 2019). This holistic perspective has only come into focus in recent years, and comparatively, there seems to be little evidence to date on the impact of AI-based decision-support systems on their users. This is likely due to the fact that in the prevailing engineering disciplines, the evaluation of systems is often a system-oriented process based on quantitative system data and does not involve users. Therefore, it is all the more important—both for current research and existing practice—to identify, summarize, and analyze previous scientific evaluation studies that include users as interaction partners of augmented intelligence systems in the evaluation, and to derive appropriate recommendations for action. In summary, this study aimed to conduct a systematic review of the literature on the impact of AI-based decision support systems on their users in a professional context.

Theoretical Background

Work-related decisions are becoming more demanding in our digitally networked working world

Decision-making is an integral part of employees' everyday work. Classic examples comprise physicians making diagnosis- and treatment-related decisions, bankers making investment decisions, or project developers deciding on numerous aspects in the course of a construction project, for example, statics, insulation, and materials. These examples demonstrate that the circumstances of work-related decisions are very different in nature; however, the common element in each situation is that there are several alternative courses of action to choose from. As the final choice should be made consciously and voluntarily, alternative courses of action should be compared, weighted, and evaluated (Rau et al., 2021). Therefore, decision-making can be understood in terms of the weighing process (Büssing & Glaser, 2002). In this process, it is often difficult or even impossible to consider all relevant information for decision-making. Owing to digitalization, the amount of data and information is often so voluminous that it significantly exceeds the human capacity to process and solve problems (Koltay, 2017; Saxena & Lamest, 2018; Shrivastav & Kongar, 2021). For example, consider a tax investigator. It is virtually impossible to detect fraud without technical assistance if the data is extensive. In addition to information deficits, employees are often confronted with further stress factors such as task complexity and time pressure (Phillips-Wren & Adya, 2020). Furthermore, in several cases, the cognitive demands placed on professionals have been increasing (Zhang et al., 2019); for example, consider physicians, who constantly have to master new and more complex diagnostic and treatment methods.

When the cognitive demands of a work task or work-related decision situation exceed our individual resources (e.g., cognitive capacity or expertise)—and this imbalance is also possibly exacerbated by work stressors such as time constraints—we are likely to end up in a state of overload, which manifests itself in high mental effort or immediate frustration and stress. The experience of psychological overload in a decision-making situation can have short-, medium-, and long-term consequences. In the short term, feelings of dissatisfaction with the decisions made or feelings of exhaustion are likely. Medium-to long-term job dissatisfaction may also possibly occur. In contrast, where there is a balance between work demands and individual resources, employees feel activated in the decision-making situation. This, in turn, can be reflected in increased performance and work engagement. Unlike in its colloquial use, the term psychological "workload" in ergonomic sciences is value-neutral and, thus, neither positive nor negative (Lazarus & Folkman, 1984; Rohmert, 1984; Karasek, 1979; Wieland & Hammes, 2014; see Figure 1).

[Figure 1 near here]

AI-based decision support systems as reducers of psychological overload

To mitigate the unfavorable effects of complex decision-making situations on employees' psychological workloads during task processing and the associated undesirable consequences (e.g., dissatisfaction and low performance), organizations are increasingly relying on the use of technical solutions in professional contexts (Tortorella et al., 2022). As described in the Introduction, there is much potential in providing employees with decision support systems (DSSs) based on AI methods as an aid (Kim et al., 2022). In general, DSSs are computer-based tools that increase users' effectiveness in making complex decisions and in accessing accurate information on available decision alternatives (Power et al., 2011). Three characteristics of DSSs are included in Power et al.'s (2011) definition. First, these systems are computer-based and, therefore, are developed in the form of software. Second, they must be effective in improving the quality of decision-making. Therefore, the outcome achieved with DSSs should be better than that achieved without them. Third, DSSs are created for complex decisions, for example, in situations where there are many decisive factors or many dynamics and uncertainties. For this purpose, systems must contain a model representing various decision elements, such as objectives, decision factors, linkages, and restrictions. Systems that do not aim to improve decision quality (and instead, for example, aim to increase efficiency) or that do not contain a decision model (and instead, for example, process information in a decision-oriented manner) are often referred to as DSSs in a broader sense (Lausberg & Krieger, 2021).

Therefore, what is special about DSSs based on AI methods? First, it is important to consider that even though the term “artificial intelligence” is omnipresent, there is no universally accepted definition of it (Duan et al., 2019; Wang, 2019). However, there is agreement that AI “is a broad umbrella term used to encompass a wide variety of subfields dedicated to, simply put, creating algorithms to perform tasks that mimic human intelligence.” (Mutasa et al., 2020, p. 96). Today, DSSs based on AI often rely on the subfield of machine learning (ML) (Janiesch et al., 2021). ML deals with computer methods that automatically learn and improve from experience or data, without being explicitly programmed. Therefore, systems based on the ML approach do not operate based on rules coded by humans, as is the case with expert systems. ML algorithms instead have the ability to create their program code independently as they automatically develop their own model from existing data and, thus, define their own rules. They do this by identifying patterns and relationships in underlying data. The learned models are then applied to new, unknown datasets of the same type to make predictions (Murphy, 2012; Janiesch et al., 2021). This method allows access to several new fields of application and the development of particularly powerful systems. A classic example is an AI-based DSS in the medical field that learns from previous patient data, such as image data, and makes diagnostic recommendations.

Such AI-based systems have already been in practice, and there are even a few reports on the impact of AI-based DSSs on employees. For instance, a field study shows that three months after the introduction of an AI-based DSS, the users’—in this case, nurses—work satisfaction concerning their diagnoses increased from 41.1% to 75%. Further, the overall time spent on diagnostic decision-making was reduced from 35.5 to

19.8 minutes (Liao et al., 2015). Tang et al. (2019) demonstrated similar efficiency. Their pilot study in a nursing home showed that the time for formulating a nursing care plan for elderly patients could be reduced from seven days to four days by introducing an AI-based DSS. However, at this point, it should be noted that the introduction of the AI-based system was accompanied by process adjustments and digitalization activities, which could potentially explain these effects. In this case, the efficiency of care plan creation largely depends on the waiting time for important documents. During the study period, admission staff contacted 18 applicants to obtain supplementary documents, and all applicants were able to upload the relevant documents within six hours. Compared to the traditional approach, in which the waiting time for supplementary documents amounted up to 24h, a reduction of 75% was achieved. This example clearly shows that the side effects of system implementation projects, such as digitalization or other organizational reasons, could potentially explain these effects. This raises the following question:

Research Question (RQ) 1: What direct influence does the use of AI-based DSSs in work-related decision-making situations have on users' psychological workload and what are the related consequences?

AI-based decision support systems can have different characteristics

It may be assumed that the effects of AI-based DSS outcomes specifically depend on how the systems are designed in everyday professional life. In particular, they have special characteristics compared to classic information systems (Shin, 2020). Two of these—transparency and accuracy—are briefly discussed below as examples.

As AI-based systems often develop their own programming rules, the algorithmic mechanisms for model generation are, in many cases, not transparent, or at least, less transparent than the logic of conventional rule-based DSSs (Jussupow et al., 2022). Therefore, the underlying logic of these systems is often referred to as the black box model (Kraus et al., 2021). The lack of information about why an AI-based system functions in certain ways can have an impact on users' trust and acceptance of AI-based DSSs (Shin et al., 2020). Moreover, the transparency users perceive in AI-based DSSs could also influence their evaluation capability regarding the system output; this plays an important role in the decision-making process. This is because the key to success in AI-assisted decision-making lies in the formation of a correct mental model of the functioning of the AI-based system and, thus, its error margins. This means that decision makers need to recognize when to trust or distrust the model's advice. If they mistakenly follow the model's advice when it is likely to be flawed, the outcome of the decision would suffer, and fatal errors could occur (Zhang et al., 2020). In particular, this applies to areas of work where decisions often lead to serious consequences, such as criminal law or healthcare. In addition, a branch of research in explainable artificial intelligence (XAI) is trying to develop tools that help understand how an AI-based system arrives at its output and what goes into this output (see Arrieta et al., 2020).

The descriptions above have addressed another peculiarity of AI-based DSSs: Algorithm performance can be impressive, but due to their probabilistic nature, there is no guarantee of the correctness of a given result (Zhang et al., 2020). Therefore, users must deal with systems that ideally have a high probability of providing correct outputs but can still be incorrect. Consequently, many companies inform their users about AI-based system performance metrics in practice, such as accuracy rates, so that they can better assess the system's performance. Depending on how well an AI-based system performs, a user perceives the system as more or less trustworthy (Papenmeier et al., 2019). It can be assumed that users are more likely to follow the advice of a system and are less likely to question it if they perceive it to be particularly accurate and, therefore, trustworthy. This possibly leads to a greater reliance on the system and, consequently, on cognitive relief, which presumably lowers the experience of stress in decision-making.

The described characteristics clarify that AI-based DSSs differ from conventional information systems or DSSs in several ways. As such, it is important to determine how the different system characteristics of AI-based DSSs affect the user; therefore, we formulated the following research question:

RQ 2: How do individual characteristics of AI-based DSSs in work-related decision-making situations affect users' psychological workload and what are the related consequences?

Methodology

To answer the aforementioned research questions, we conducted a systematic literature review wherein we summarized the results of existing studies on the impact of AI-based DSSs on users' psychological workload and related consequences. We adhered to the Preferred Reporting Item for Systematic reviews and Meta-Analysis (PRISMA) statement guidelines (Page et al., 2021). Following the PRISMA flowchart, this section describes the methodological approach in three steps: (1) identifying relevant studies, (2) selecting studies, and (3) analyzing the included studies and synthesizing findings. The first two steps are described in this section; the third is discussed in the Results section.

Identifying relevant studies

To identify and extract scientific journal articles addressing the relationship between AI-based decision support systems and the psychological construct of psychological workload and the related consequences, we used the online psychology databases

Psyndex, PsycINFO, and PsycARTICLES from EBSCO*host*. As the use of AI-based systems in the professional context is often discussed in business and information sciences, we also conducted a literature search in the Business Source Elite and Library, Information Science & Technology Abstracts. To search the database, we first identified a set of keywords related to our research topics in English and German as we considered scientific works in both languages in this study. Two groups of keywords were distinguished: The first group included keywords for systems based on AI methods, and the second group related to the scope of application of these systems, that is, the professional context and their impact on users. In detail the applied keywords were: "artificial intelligence*," "augmented intelligence*," "intelligence augmentation*," "AI*," "decision support*," "decision making*," "AI assistance system*," "humancentred artificial intelligence*," "industry 4.0*," "human machine interaction*," "human machine interface*," "work*," "mental*," "job*," "psychological*," "working conditions*," "Künstliche Intelligenz*," "KI-System*," "KI-Assistenzsysteme*," "KI-basierte Systeme*," "Industrie 4.0*," "KI Mensch-Schnittstelle*," "Mensch-Maschine-Schnittstelle*," "Menschzentrierte Künstliche Intelligenz*," "Arbeit*," "Belastungen*," "Beanspruchungen*," "Psyche*," "Psychische*". We ran a query using a combination of these keywords (adopting the Boolean operator "AND"). We employed this wide approach to ensure that we could find as many relevant studies as possible.

A literature search was conducted in January 2022 and completed in December 2022. In January 2022, the aforementioned search strategy resulted in 10,367 relevant articles after using the filter function in the databases to exclude papers that were not published in academic journals, English, German, or that were published before 2012. The full inclusion and exclusion criteria are presented in Table 1. We focused on publications from 2012 onwards because there have been significant advances in the field of AI in recent years and, thus, in the specific functionalities and characteristics of corresponding applications (Nicodeme, 2020). Moreover, in recent years, the topic has been increasingly viewed from a transdisciplinary perspective (for example, see Y. Cai et al., 2019). Following the database search in January 2022, an exploratory search using a snowball system (Wohlin, 2014) for suitable articles was conducted, resulting in 42 articles. A second search was conducted in September 2022 to identify recent publications from 2022 and yielded another 918 results. Another manual search yielded 15 additional relevant studies. The final search was conducted in December 2022 to include all publications from 2022. In total, 236 studies were included. Consequently, 11,578 papers were found, and after excluding duplicate papers, 9,908 papers were retained.

[Table 1 near here]

Study selection

The study selection was performed in two screening steps. First, the titles, abstracts, and keywords were checked, and unsuitable studies were eliminated. Subsequently, the remaining studies were read completely and classified into “include,” “exclude” or “maybe” categories. Two independent reviewers conducted both steps. For this process, we used the free platform Rayyan (https://www.rayyan.ai/). Thereafter, the reviewers discussed those studies that were assigned to the category “maybe” or to the “include” category by only one reviewer. In cases wherein disagreements persisted, a third reviewer was consulted. Finally, consensus was reached in all cases.

The entire selection process is illustrated in Figure 2 in the form of a flow diagram. Following the title, abstract, and keyword screening, 9,693 irrelevant documents were excluded. The full texts of the remaining 215 articles were then examined. In this step, 199 studies were further excluded. Most of these articles were excluded in the first stage of full-text screening, primarily because they did not focus on AI-based DSSs, but instead, on AI in general or on fully automated AI-based systems. Another frequent reason for exclusion was that studies only evaluated the system, for example, how well it worked or how trustworthy it was, but not its impact on users. Another reason was that the studies dealt exclusively with augmented intelligence systems theoretically. Moreover, one study (Calisto et al. 2021) was replaced with an updated and republished version (Calisto et al. 2022). Ultimately, 16 journal articles were selected for this review.

[Figure 2 near here]

Results

Descriptive analysis of the included studies

This review included 15 primary studies and 1 secondary study, more specifically, a systematic review that also summarized experimental studies. All studies (see Table 2) were published between 2016 and 2022, and most of them were published in 2021 and 2022. Moreover, all studies were peer-reviewed, with the exception of Gaube et al. (2022), whose work is a preprint and is currently in the peer review stage. Most of the included studies (*n*=10) addressed the use of AI-based DSSs in the field of medicine, followed by three studies in the military domain. Most studies used samples of task experts; therefore, with a focus on medicine, most participants were physicians. The remaining studies that used lay people as participants asked them to imagine being an employee during the experimental task, for example, a hiring manager in human resources (see Table 2). A total of 1,332 people participated in the 15 primary studies, with cohort sizes varying from *N*=5 (M. H. Lee et al., 2021) to *N*=265 (Gaube et al., 2021). From the systematic review by Li et al. (2021), the pooled results of 119 participants from a total of fourteen investigations were reported (Bai et al., 2021; Dorr et al., 2020; Kim et al., 2020; Koo et al., 2020; Kozuka et al., 2020; Liu et al., 2019; K. H. Lee et al., 2012; Martini et al., 2021; Nam et al., 2021; Rajpurkar et al., 2020; Singh et al., 2021; Sung et al., 2021; Yang et al., 2021; Zhang et al., 2021). In these investigations, the cohort size varied from N=2 (Kozuka et al., 2020; Liu at al., 2019; Singh et al., 2021; Zhang et al., 2021) to N=54 (Dorr et al., 2020). The research design of all studies—including those from the secondary study by Li et al. (2021)—was in almost all cases an experimental setting with a randomized control group design (Brauner et al. 2019; C. J. Cai et al., 2019; Calisto et al., 2022; Dorr et al., 2020; Didimo et al., 2018; Gaube et al., 2022; Gaube et al., 2021; Jacobs et al., 2021; Jussupow et al., 2022; Kim et al., 2020; Knapič et al., 2021; Koo et al., 2020; Kozuka et al., 2020; Langer et al., 2021; K. H. Lee et al., 2012; M. H. Lee et al., 2021; Liu et al., 2019; Martini et al., 2021; Nam et al., 2021; Rajpurkar et al., 2020; Roth et al., 2020; Singh et al., 2021; Stowers et al., 2020; Sung et al., 2021; Vodrahalli et al., 2022; Zhang et al., 2021). We could not confirm this statement for three studies as they did not include information on whether randomization took place (Bai et al., 2021; Mercado et al., 2016; Yang et al., 2021).

Of the 16 studies included, 8 studies investigated the overall effect of AI-based DSSs in the professional context on their users and, thus, contributed to answering the first research question. Of these, six studies specifically investigated this effect (C. J. Cai et al., 2019; Calisto et al., 2022; Didimo et al., 2018; Langer et al., 2021; Lee M. H. et al., 2021; Li et al., 2021), and two studies observed the effect at the margin (Jacobs et al., 2021; Vodrahalli et al., 2022). In their experiments, all studies compared participants who were supported by an AI-based DSS in completing a simulated work task with those who were not supported by an AI-based DSS in completing the task. To achieve this, some of the studies used existing AI-based DSSs in their experiments, and some simulated AI-based DSSs completely or partially for their experiments. In their experiments, half of the included primary studies used a within-subject design, and the other half used a between-subject design (see Table 2). In the secondary review by Li et al. (2021), all relevant investigations listed here also used a within-subject design (Bai et al., 2021; Dorr et al., 2020; Kim et al., 2020; Koo et al., 2020; Kozuka et al., 2020; K. H. Lee et al., 2012; Liu et al., 2019; Martini et al., 2021; Nam et al., 2021; Rajpurkar et al., 2020; Singh et al., 2021; Sung et al., 2021; Yang et al., 2021), with the exception of Zhang et al. (2021), which used a between-subjects design.

Of the 16 included studies, 11 studies examined the effects of individual characteristics of AI-based DSSs and, thus, contributed to answering the second research question. Specifically, the included studies examined the role of *correctness, transparency,* and *timing of support* in AI-based DSSs. Eight studies were fully simulated, one study by Vodrahalli et al. (2022) partially simulated an AI-based DSS, and two studies evaluated a real AI-based DSS. In addition, 6 of the 11 studies used a within-subjects design in their experiment, 2 used a between-subjects design, and 3 used a 2×2 factorial design (see Table 2).

[Table 2 near here]

Results for RQ1: General impact of AI-based DSS on employees

The first research question addressed the general impact of the use of AI-based DSSs in work-related decision-making situations on users' psychological workload and the related consequences. Among the included studies, only two studies by C. J. Cai et al. (2019) and M. H. Lee et al. (2021) investigated the impact on users' immediate psychological workload during task processing. Both studies show that employees who are supported in task processing by an AI-based DSS exert significantly less mental effort than those supported by a traditional information system. However, no significant differences were found in either study regarding participants' experience of frustration during task processing.

In addition to these studies, six other studies examined the general effect of AI-based DSSs on the workload consequences of users, more specifically, their task performance (see Table 2). This number included the systematic review by Li et al. (2021): Their secondary study predominantly reported research findings revealing that physicians supported by AI-based DSSs in the detection of pathologies make significantly better diagnostic decisions than when they do not receive such support. This is measured by diagnostic sensitivity and/or specificity (Bai et al., 2021; Dorr et al., 2020; Kim et al., 2020; Kozuka et al., 2020; Sung et al., 2021; Yang et al., 2021), diagnostic accuracy (Bai et al., 2021; Nam et al., 2021; Rajpurkar et al., 2020; Yang et al., 2021), and area under the receiver operating characteristic curve (AUC) (Kim et al., 2020; Koo et al., 2020; Sung et al., 2021).

The review by Li et al. (2021) also reported five studies describing that the use of AI-based DSSs leads to improved diagnostic performance; however, they did not report whether this effect was significant (Koo et al., 2020; M. H. Lee et al., 2021; Liu et al., 2019; Rajpurkar et al., 2020; Singh et al., 2021). In addition, three of the five primary studies reported that AI-based DSSs lead to performance improvements in users without reporting significant values (Calisto et al., 2022; Didimo et al., 2018; Vodrahalli et al., 2022). For example, Didimo et al. (2018) evaluated the impact of a real AI-based DSS—in this case, a system called TAXNET. The system was designed to help tax investigators detect tax evasion more easily. The experiment showed that participants, employees of the Italian tax authority using TAXNET, made better decisions in both experimental tasks when supported by an AI-based DSS (task 1: 98.98% accuracy rate; task 2: 98.83% accuracy rate) than when supported by a traditional information system (task 1: 87.08% accuracy rate; task 2: 63.09% accuracy rate). Meanwhile, Vordrahlli et al. (2022) did not observe such high levels of performance increases in their study. However, the authors reported performance improvement in the form of an increase of 6.3% in the accuracy rate with AI-assisted support, compared to without AI-assisted support.

Didimo et al. (2018) also observed that on average, participants using TAXNET needed only approximately 52% of the time required by participants without TAXNET to complete the first task. Calisto et al. (2022) and seven of the studies included in the review by Li et al. (2019) also reported time savings from AI support (Kim et al., 2020; Kozuka et al. 2020; Liu et al., 2019; Martini et al., 2021; Nam et al., 2021; Sung et al., 2021; Zhang et al., 2021); however, only three studies indicated that the difference was significant (Martini et al., 2021; Nam et al., 2021; Sung et al., 2021).

Moreover, the review by Li et al. (2021) reported a study by Singh et al. (2021) that found no significant difference in task performance in terms of the AUC performance measure between the two experimental conditions (with vs. without AI-based DSS). Further, two primary studies included in this review also did not find any significant difference in task performance between individuals who used an AI-based DSS and those who did not (Jacobs et al., 2021; Langer et al., 2021).

Results for RQ2: Impact of individual characteristics of AI-based DSSs on employees

Correctness of AI-based system support

Four papers investigated what single influences the correctness of an AI-based system output or advice has on its users (see Table 3). All four studies showed that this characteristic has a central influence on users' task performance. The majority of the results showed that incorrect AI-based advice is associated with significantly lower performance scores among users compared to correct AI-based advice (Gaube et al., 2021; Jacobs et al., 2021; Jussupow et al., 2022). This also applies in the comparison of the baseline condition, in which participants do not receive AI-based system support (Jacobs et al., 2021; Jussupow et al., 2022). It is unsurprising that incorrect system advice leads to worse performance compared to correct system advice. However, what is surprising is that in two studies, no significant differences in objective performance measures were observed when subjects received correct AI-based advice compared to no AI-based system support on a work task (Jacobs et al., 2021; Jussupow et al., 2022).

Moreover, one study by Brauner et al. (2019) showed that participants' self-perception of their performance compared to others was also influenced by the correctness of the system advice. In their experiment, participants went through two rounds of a business simulation game in the role of supply chain manager. The AI-based system worked correctly in one round and incorrectly in another. The assignments were made randomly. In detail, the results showed that the subjects who used the defective system in the first round reported a 38% lower relative self-perception of their performance than those participants who used the correct system ($F_{1, 38}$= 4.11, p= .050). In the second round of experiments, the subjects who were now given an incorrect system had a 13% lower relative self-assessment of their performance compared to those who used the correct system in this round ($F_{1, 38}$= 1.18, p= .285). The same pattern was observed for respondents’ satisfaction with their own performance. In the first round, participants’ satisfaction with the incorrect DSS was 46% lower than that of participants with the correct system ($F_{1, 38}$= 5.01, p= .031). In the second round, participants’ satisfaction with the incorrect DSS was 19% lower compared to that of participants with the correct system ($F_{1, 38}$= 2.14, p= .152), though not significantly.

Jacobs et al. (2021) and Gaube et al. (2021) examined in their studies the effect of the correctness of a system’s advice on physicians’ own confidence in a decision taken. These studies observed different effects. Jacobs et al. (2021) found no main effect of correctness of advice on physicians’ confidence in medical treatment decisions between conditions without system support (M= 3.67, 95% CI [3.63; 3.72]), correct AI-based advice (M= 3.65, 95% CI [3.62; 3.69]), and incorrect AI-based advice (M= 3.62, 95% CI [3.57; 3.67]; $F_{2, 3379}$ = 2.02, p= .133). In contrast, Gaube et al. (2021) observed that all participants—including both high and low level task experts—were significantly more confident in their diagnosis when the advice was accurate (high-level experts in the form of radiologists: $t(137)$= 6.65, p< .001; low-level experts in the form of physicians trained in internal/emergency medicine: $t(126)$= 8.43, p< .001).

[Table 3 near here]

Transparency of AI-based system support

Of the included studies, six dealt with the question of how the transparency level of an AI-based DSS, in the form of different levels of additional information/explanations on the system output, affects users (Gaube et al., 2022; Jacobs et al., 2021; Knapič et al., 2021; Mercado et al., 2016; Roth et al., 2020; Stowers et al., 2020). Two studies examined the effects of specific XAI methods on users (Jacobs et al., 2021; Knapič et al. 2021). As we were interested in the general system characteristics from which general design recommendations can be derived but not in the comparison of specific XAI methods, these two papers were not considered further. The remaining studies, except one, showed that additional information on/explanations of a system's output have a significant influence on users' task performance (see Table 4). For example, Gaube et al. (2022) found that physicians made better diagnostic decisions when given AI-based advice with a visual explanation than when given an AI-based advice with no visual explanation (*OR*= 2.30, *SE*=0.61, 95% CI [1.37; 3.86], z=3.17, p<.002). Looking at the results separately for different target groups—task experts in the form of radiologists and non-task experts in the form of physicians trained in internal medicine or emergency medicine (IM/EM)—the authors revealed that annotated advice significantly improved the performance of IM/EM physicians but had no significant effect on the diagnostic accuracy of radiologists. Throughout the experiment, the performance was high, although it was significantly lower for one case under both advice conditions. Interestingly, the annotations in this difficult case also seemed to have a positive effect on task experts.

Furthermore, Mercado et al. (2014) and Stowers et al. (2020) highlighted that additional information on/explanations of the system output significantly assist users in identifying correct system advice. Mercado et al. (2016) also revealed a significant main effect of additional information on/explanations of system output on users' performance in recognizing and rejecting incorrect advice ($F_{2, 58}$= 12.33, p< .001). Replicating the Mercardo et al.'s study (2016), Stowers et al. (2020) did not observe this effect in their work. Roth et al. (2020) also reported a noticeable but non-significant effect. The results of the study by Roth et al. (2020) showed no significant effect of different levels of information on/explanations of system output on user performance in terms of quality but did reveal a significant effect in terms of time: The results demonstrated that the time it took to determine the validity of AI-based advice was significantly reduced by providing additional information/explanations. In contrast, Mercado et al. (2016) found no significant difference in response times between different levels of information/explanations. The replication study by Stowers et al. (2020), in turn, showed a significant main effect of additional information on/explanations of system output on response time ($F_{2,\,104}$= 4.37, *p*= .015, $\eta 2$= 0.02), but in contrast to Roth et al. (2020), the significant results demonstrated that the reaction time increased significantly in the high transparency condition (*M*= 39.47s, *SD*=16.6s) compared to the low transparency condition (*M*= 33.65s, *SD*= 19.68s).

However, regarding the influence of the system transparency level in the form of additional information on and explanations of system output on users' immediate psychological workload experience during task processing, the studies showed consistent results. Mercado et al. (2016), Stowers et al. (2020), and Roth et al. (2020) did not find any influence of the system's transparency level on users' subjective workload during task processing. Mercado et al. (2016) also revealed no effect on objectively measured workloads using eye-tracking. Furthermore, Gaube et al. (2022) identified no difference in the assessment of the decision confidence of subjects with different levels of information on and explanations of system output (*b*=0.06, *SE*=.07, 95% CI [-0.09; -0.20], *t*(222)=0.77, *p*=.440).

[Table 4 near here]

Timing with AI-based system support

Two of the included studies addressed the following question: Where in the decision-making process should an AI-based DSS offer its support (Jussupow et al., 2022;

Langer et al., 2021)? To do so, they compared the use of an AI-based DSS at two different positions in the decision-making process: once before a human decision-maker processed the available information for decision-making (support-before-processing), and once after the initial processing of information for decision-making had taken place (support-after-processing). Langer et al. (2021) hypothesized that individuals in the support-before-processing condition would complete the experimental task in less time but would also enjoy the task less and perceive it as more monotonous than individuals in the support-after-processing condition or individuals who received no AI-based system support at all. In addition, the authors hypothesized that individuals who received support-before-processing would feel less responsible for their decisions than those who received support-after-processing or no system support. Nonetheless, none of these hypotheses could be confirmed in their study. However, the authors found evidence that participants in the support-after-processing condition were more satisfied with their decisions than participants in the combined sample of the support-before-processing condition and no AI-based support condition ($t(119)= 2.47$, $p= .01$, d=0.46). They also observed that individuals in the support-after-processing condition showed significantly higher general self-efficacy ($t(119)= 2.74$, $p= .01$, $d=0.51$) and specific self-efficacy ($t(119)= 2.53$, $p= .01$, $d=0.49$) scores than the other two groups. In addition, participants in the support-after-processing condition reported that they enjoyed the task slightly more than the no-support group ($t(84)= 1.80$, $p= .04$, $d=0.40$). However, another hypothesis was that participants in the support-after-processing condition experienced less monotony in task completion than participants without AI-based support ($t(84)= 1.19$, $p= .12$, $d=0.25$). In addition, Jussupow et al. (2022) found that the timing of the AI-based system support had no influence on users' performance.

Discussion

Augmented intelligence, as a design approach in the field of human-computer interaction, is gaining increasing popularity, both in practice and research. Similar to human-computer interaction, this approach aims to use machine intelligence to improve users' experiences, task performance, and quality of life (Abbassi et al., 2020; Kim et al., 2022). This review explores whether and under what conditions these desired effects on users actually occur in the professional context. For this purpose, a systematic literature search was first conducted, resulting in the inclusion of a total of 16 experimental studies that investigated the effects of AI-based DSSs on users in the professional context.

Overall, the relatively small number of identified studies shows that research examining the actual impact of AI-based DSSs on its users is still in its early stages. Fortunately, the results also indicate that research interest in this topic has been growing significantly in recent years (see Table 2). However, most studies have focused on the impact of AI-based DSSs on the consequences of psychological workload for employees—primarily in terms of performance—and only a few studies have examined the direct workload experience of users in decision-making situations. In line with this, only two of the identified studies examined the overall effect of AI-based DSSs on users' psychological workload during decision-making, providing evidence that employees in work-related decision-making situations exert less mental effort with AI support than without (C. J. Cai et al., 2019; M. H. Lee et al., 2021). The remaining papers investigating the impact of AI-based DSSs on users' task performance also revealed promising results. Most studies found that AI-based DSSs lead to improved performance (Li et al., 2021). However, some studies do not state whether these improvements are statistically significant (e.g., Calisto et al., 2022; Didimo et al., 2018). Furthermore, there are single studies that report no significant improvements in performance (Jacobs et al., 2021; Langer et al., 2021). These results support the assumption that—in accordance with existing stress theories (for example, see Rohmert, 1984; Karasek, 1979)—situational factors also affect the effects of AI-based DSSs on users. Such situational factors can be requirements in a work-related decision situation, the individual characteristics of an AI-based DSS as an organizational resource, or the personal characteristics of the user as individual resources (see Figure 1).

It is not surprising that the impact of an AI-based DSS on its users depends—among other factors—on the nature of the decision-making situation, especially on its complexity. The study by Langer et al. (2021) illustrates this very well: In their experiment, lay people were asked to put themselves in the role of a hiring manager and make personnel decisions; half of the subjects were supported by an AI-based DSS, while the other half were not. The authors found no significant difference in performance between the two groups. However, they noted a lack of variance in the performance of the participants: Performance was very high in all cases—with or without system support. This means that if a work task or decision situation is not particularly complex and can be solved easily without system support, an AI-based DSS provides relatively less advantage. The fact that the usefulness of a system depends on the difficulty of the decision task is also evident in two other studies included in our review (Calisto et al., 2022; Didimo et al., 2018). In these studies, the participants solved tasks with varying degrees of difficulty. Under the most difficult condition, the difference in performance with and without AI-based support was the greatest. As previously described, these results are unsurprising, though they underline that AI-based DSSs are developed especially for complex decision-making situations and show their benefit here. In addition, less complex use cases are likely to run the risk that employees will be challenged inadequately by system support during task processing, leading to monotony and fatigue.

The summarized results of the review also indicate that the impact of AI-based DSSs on their users depends on how detailed the application design is. The results of these studies particularly emphasize the role of accuracy in an AI-based DSS. If accuracy is not sufficiently high, as was the case in the study by Jacobs et al. (2021), the system does not have a significantly positive impact on the performance of users; on the contrary, it can even worsen their performance. More specifically, the study results demonstrate that users' performance is worse when they receive incorrect advice compared to correct or no advice. Accordingly, it can also be observed that subjects in these cases feel significantly less confident in the decisions they make and are significantly less satisfied with them (Brauner et al., 2019; Gaube et al., 2021). In contrast, Jacobs et al. (2019) found no significant differences between the groups—correct vs. incorrect AI-based decision advice—in terms of users' perceived confidence in the decisions they made. However, we believe that these results can be explained by the experimental decision situation. In their experiment, subjects were asked to choose a treatment method for a patient with a major depressive disorder. The peculiarity of this situation is that there is no single correct treatment method for these patients; instead, there are often several options to choose from. It is possible that the subjects—in this case, experienced physicians as task experts—are already instructed that there are different possible treatment options and are, therefore, less irritated by other suggestions, even if they are wrong. However, overall, it can be deduced from the summarized study results that users are negatively influenced by inaccurate advice from AI-based DSSs.

Jussupow et al. (2022) illustrated why this aforementioned irritation is so powerful. In their study, the authors observed how participating novice physicians handled correct and incorrect AI-based system advice while making a medical diagnosis. The subjects were asked to verbalize their thoughts (think-aloud method). In this manner, the authors were able to identify the internal cognitive processes with which users processed the system's advice. From their results, the authors were able to make the following observation: When participants received system advice, they often first checked whether it matched their own mental model of a decision situation. If this was not the case, approximately half the users could be influenced and made to change their original opinion—this applied to both correct and incorrect system advice. The central point here is that the majority of participants (approximately 80%), as task experts, or possibly due to the fact that the experimental task was not very complex, had a correct mental model and made a correct initial assessment in most cases. Consequently, in the case of inaccurate system advice, the group of people who were negatively influenced by the system was much larger than the group with correct advice, as the majority was already on the right track beforehand. As a result, in the correct system condition, the overall performance improved from 80.95% to 90.48%, and in the inaccurate condition, overall performance deteriorated from 81.82% to 54.55%. This observation also illustrates why, in these cases, incorrect system advice leads to significantly worse performance than without system support, but correct advice does not lead to significantly better performance.

In conclusion, the study results indicate that many participants overtly trust AI-based system support. In this case, it could be argued that the main reason for the overestimation of the AI-based system is that the sample consisted of low-level experts with novice physicians. This seems to be true to some extent, as both Jussupow et al.

(2022) and Gaube et al. (2021) found that the tendency to follow incorrect advice was significantly higher among low-level task experts than among high-level task experts. Surprisingly, both studies also showed that the overall performance of the two groups was not significantly different. This means that highly experienced individuals are not significantly better at working with the system and distinguishing between correct and incorrect advice of AI-based DSSs. This is because they choose to ignore advice from AI-based DSSs, even if they are correct (Jussupow et al., 2022). Thus, high-level task experts are more likely to question AI-based system advice than low-level task experts (e.g., Gaube et al., 2022). This makes sense because a higher level of experience is positively correlated with people's confidence in their own decisions. The high self-confidence of high-level task experts, thus, seems to lead them to follow inaccurate AI-based system advice less but also follow correct advice less, and these effects ultimately balance out.

At this point, it should also be noted that the subjects in almost none of the experiments received any feedback on their performance or on the accuracy of the system during the experiment (see e.g., Gaube et al., 2021; Jacobs et al., 2021). Therefore, it would be interesting to determine whether users would improve their performance over time by receiving information on the overall performance of the system or feedback on their own performance. However, information about incorrect system advice could negatively affect the willingness to use AI-based systems. This consideration is supported by the results of Dietvorst et al. (2015). Considering these indications, it can be stated that only systems with a high performance level should be implemented; otherwise, there is a risk that AI-based DSSs will impede employees instead of aiding them. Alternatively, systems could operate in such a way that they only give advice if the probability of error is sufficiently low. Overall, further research is required to support these reflections.

A few of the included studies have already examined whether additional information/explanations provided by the system (i.e., the system is as transparent as possible in its decision advice) help users recognize whether a piece of advice is correct or incorrect (see Table 4). The results of two studies (Mercado et al., 2016; Stowers et al., 2020) affirmed this, although only one study (Mercado et al., 2016) yielded significant results. Roth et al. (2020) observed that participants' performance—particularly approving correct and rejecting incorrect planning proposals—showed a noticeable but not significant improvement when additional transparency information was provided. Owing to the small sample size of 10 participants, it can be assumed that this effect is not significant. Generally, the studies showed that additional information/explanation significantly improves the overall performance of users (see Table 4). The results also show that different levels of transparency have no impact on the direct psychological load of users. One possible reason is that the additional insights provide more clarity on the one hand, but on the other hand, they also require deeper processing so that the effects balance out. Therefore, it would be interesting to determine the long-term effects of different system transparency levels if users have learned to quickly assess additional information over a certain period of time.

The results of Gaube et al. (2022) also indicate a long-term effect, observing that users do not feel more certain regarding their decisions due to transparency levels. This may be because they first have to gain experience with them in order to be able to classify them. In their study, the authors also compared whether additional explanations of system output have a differential effect on task performance, as well as on non-task experts. When comparing the overall diagnostic performance of task experts (radiologists) with non-task experts (IM/EM doctors), only the latter group benefited significantly from the comments. Looking at individual patient cases, they found that task experts also benefitted from visual annotations when examining a more difficult case. Therefore, Gaube et al. (2022) suggest that task experts also benefit from annotations but that ceiling effects mask this benefit. However, further research is needed at this point.

Limitations and implications for future research

This study has several limitations. Initially, the database research only considered psychological, business administration, and information technology journals. Since the present overview has shown that there are many research efforts on this topic, especially in the medical context, we suggest continuing with future searches in medical databases. Second, our study considered only experimental studies with a control group design. This may have led to the exclusion of potentially interesting studies. Third, our study included an article by Gaube et al. (2022), which is a preprint in the review process. This should be considered when interpreting our results. Fourth, the 16 included studies examined very different types of AI-based DSSs and different subject sizes. Therefore, the results may not be generalizable, and should be interpreted with caution.

In order to be able to make more general statements in the future, further research is required. For example, it would be interesting to investigate the effect of AI-based DSSs on users' immediate workload experience in broader professional contexts, where these systems are likely to be used in the future, such as in service and customer care or the real estate industry. It would also be exciting to study mediation and moderation effects; for example, whether AI-based DSSs actually lead to fewer information deficits in decision-making situations and, thus, enable a feeling of being less overwhelmed. In addition, it would be interesting to investigate whether the introduction of these systems could also lead to unintended side effects, such as the loss of competencies, as can happen with fully automated AI-based systems (Mayer et al., 2022).

Practical implications

Despite the need for further research, our review already shows that it cannot necessarily be assumed that AI-based DSSs in the professional context have a positive influence on users. To achieve the desired effects, such as workload relief and performance increase, and avoid potential unintended side effects, such as a loss of competence, we recommend that companies systematically plan the application field, as well as the development, implementation, and evaluation of AI-based applications. An orientation for such a systematic procedure is provided, for example, by classic phase models of change management (for example, see Von Hehn et al., 2016). Following such models, the work process, in which an AI-based DSS is to be used for support, should be analyzed first. As the summarized study results underline, the use of AI-based DSSs is only suitable for work activities that have a certain degree of complexity and, therefore, require a combination of human and machine intelligence in the first place. An analysis of the work process must also consider the resources and associated competences of employees. Based on the results of this analysis, an appropriate system with specific functionalities and characteristic properties can be developed. In the case of AI-based DSSs, developers should not only pay attention to classic design criteria, such as usability, but also to characteristics such as the accuracy and transparency of the system's output. As described earlier, this can reduce the risk such that the user will either not accept the system or that their uncertainty in decision-making situations will increase; as a result, performance will deteriorate. Therefore, before a central rollout, AI-based DSSs should be carefully tested and, if necessary, improved—perhaps through interactions with future users in pilot projects. As abovementioned, we also recommend that systems be designed in a manner that they only extend advice if the probability of error is sufficiently low. If the system has a low likelihood of suggesting an accurate solution, it should instead advise the user that the best course of action is to make a decision without the DSS. In addition, AI-based DSSs output should be sufficiently transparent for users, while at the same time adhering to the design principle of economy. This means that the interface should display as much information as necessary and as little information as possible. Otherwise, there is a risk that users may be cognitively stressed by the systems, and that the desired positive relief effects will be offset. When implementing AI-based DSSs in everyday work, organizations should also consciously consider at what point in the decision-making process (once before a human decision maker processes available information for decision making or after some initial processing of information for decision making has taken place) they are deploying AI-based systems and how that may impact users.

In summary, we recommend that companies focus on users throughout the change process. In other words, a human-centric approach should be followed not only in the development and implementation of AI-based DSSs but also consistently in the final step of evaluation. This means that the evaluation should not only consider technical assessment criteria such as the accuracy and precision rates of AI-based DSSs, but also the users: how they perceive the system, for example, as trustworthy, transparent, and usable, and what actual impact the system has on the target group, for example, on their mental effort in the decision-making situation, their feelings of confidence in the decisions they make, and their performance.

Funding

The research presented in this paper has been carried out within the research project "AIXPERIMENTATIONLAB" (Project number EXP.01.00016.20). The authors gratefully acknowledge the support of the German Federal Ministry of Labour and Social Affairs (BMAS).

Disclosure statement

The authors report there are no competing interests to declare.

References

Abbasi, A., Robert, L., Wang, W., & Wu, L. (2020). *Human Computer Interaction, Artificial Intelligence and Intelligent Augmentation.* Aisle. https://aisel.aisnet.org/icis2020/hci_artintel/

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

Bai, H. X., Wang, R., Xiong, Z., Hsieh, B., Chang, K., Halsey, K., Tran, T. M. L., Choi, D., Shi, L. B., Mei, J., Jiang, X., Pan, I., Zeng, Q., Hu, P., Li, Y., Fu, F., Huang, R. Y., Sebro, R., … Liao, W. H. (2020). AI augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT. *Radiology*, *296*(3), 156–165. https://doi.org/10.1148/radiol.2020201491

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial Intelligence in organizations: Current state and future opportunities. *MIS Quarterly Executive*, *19*(4), 9–21. https://doi.org/10.2139/ssrn.3741983

Brauner, P., Philipsen, R., Calero Valdez, Z., & Ziefle, M. (2019). What happens when decision support systems fail? - The importance of usability on performance in erroneous systems. *Behaviour & Information Technology*, *38*(12), 1225–1242. https://doi.org/10.1080/0144929X.2019.1581258

Büssing, A., Glaser, J., & Höge, T. (2004). Psychische und physische Belastungen in der ambulanten Pflege: Ein Screening zum Arbeits- und Gesundheitsschutz. *Zeitschrift für Arbeits- und Organisationspsychologie A&O*, *48*(4), 165–180. https://doi.org/10.1026/0932-4089.48.4.165

Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., Wattenberg, M., Viegas, F., Corrado, G. S., Stumpe, M. C., & Terry, M. (2019). Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In *CHI 2019: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). ACM. https://doi.org/10.1145/3290605.3300234

Cai, Y., Ramis Ferrer, B., & Luis Martinez Lastra, J. (2019). Building university-industry co-innovation networks in transnational innovation ecosystems: Towards a transdisciplinary approach of integrating social sciences and artificial intelligence. *Sustainability*, *11*(17), 4633. https://doi.org/10.3390/su11174633

Calisto, F. M., Santiago, C., Nunes, N., & Nascimento, J. C. (2021). Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification. *International Journal of Human-Computer Studies, 150*, 102607. https://doi.org/10.1016/j.ijhcs.2021.102607

Calisto, F. M., Santiago, C., Nunes, N., & Nascimento, J. C. (2022). BreastScreening-AI: Evaluating medical intelligent agents for human-AI interactions. *Artificial Intelligence in Medicine*, *127*, 102285. https://doi.org/10.1016/j.artmed.2022.102285

Didimo, W., Giamminonni, L., Liotta, G., Montecchiani, F., & Pagliuca, D. (2018). A visual analytics system to support tax evasion discovery. *Decision Support Systems*, *110*, 71–83. https://doi.org/10.1016/j.dss.2018.03.008

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

Dorr, F., Chaves, H., Serra, M. M., Ramirez, A., Costa, M. E., Seia, J., Cejas, C., Castro, M., Eyheremendy, E., Slezak, D. F., & Farez, M. F. (2020). COVID-19 pneumonia accurately detected on chest radiographs with artificial intelligence. *Intelligence-Based Medicine*, *34*, 100014. https://doi.org/10.1016/j.ibmed.2020.100014

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Gaube, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lermer, E., Coughlin, J. F., Guttag, J. V., Colak, E., & Ghassemi, M. (2021). Do as AI say: Susceptibility in deployment of clinical decision-aids. *Npj Digital Medicine, 4*, 31. https://doi.org/10.1038/s41746-021-00385-9

Gaube, S., Suresh, H., Raue, M., Lermer, E., Koch, T., Hudecek, M., Ackery, A. D., Grover, S. C., Coughlin, J. F., Frey, D., Kitamura, F., Ghassemi, M., & Colak, E. (2022). Who should do as AI say? Only non-task expert physicians benefit from correct explainable AI advice. *Research Square Platform LLC*. Advance online publication. https://doi.org/10.21203/rs.3.rs-1687219/v1

Geyer, W., Weisz, J., Pinhanez, C. S., & Daly, E. (2022, March 8). *What is human-centered AI?.* IBM. https://research.ibm.com/blog/what-is-human-centered-ai

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future? *AI*, *1*(2), 143–155. https://doi.org/10.3390/ai1020008

Hellebrandt, T., Huebser, L., Adam, T., Heine, I., & Schmitt, R. H. (2021). Augmented Intelligence – Mensch trifft Künstliche Intelligenz. *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, *116*(6), 433–437. https://doi.org/10.1515/zwf-2021-0104

Jacobs, M., Pradier, M. F., McCoy Jr., T. H., Perlis, R. H., Doshi-Velez, F., & Gajos, K. Z. (2021). How machine-learning recommendations influence clinician treatment selections: The example of antidepressant selection. *Translational Psychiatry*, *11*, 108. https://doi.org/10.1038/s41398-021-01224-x

Jain, H., Padmanabhan, B., Pavlou, P. A. & Raghu, T. S. (2021). Editorial for the Special Section on Humans, Algorithms, and Augmented Intelligence: The Future of Work, Organizations, and Society. *Information Systems Research*, *32*(3), 675–687. https://doi.org/10.1287/isre.2021.1046

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence. *Information Systems Research*, *32*(3), 713–735. https://doi.org/10.1287/isre.2020.0980

Karasek, R. A. (1979). Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign. *Administrative Science Quarterly*, *24*(2), 285–308. https://doi.org/10.2307/2392498

Kim, J., Davis, T., & Hong, L. (2022). Augmented Intelligence: Enhancing Human Decision Making. In M. V. Albert, L. Lin, M. J. Spector, L. S. Dunn (Eds.), *Bridging Human Intelligence and Artificial Intelligence* (pp. 151–170). Springer Cham. https://doi.org/10.1007/978-3-030-84729-6_10

Kim, J. H., Kim, J. Y., Kim, G. H., Kang, D., Kim, I. J., Seo, J., Andrews, J. R., & Park, C. M. (2020). Clinical Validation of a Deep Learning Algorithm for Detection of Pneumonia on Chest Radiographs in Emergency Department Patients with Acute Febrile Respiratory Illness. *Journal of Clinical Medicine, 9*(6), 1981. https://doi.org/10.3390/jcm9061981

Kirste, M. (2019). Augmented Intelligence – Wie Menschen mit KI zusammenarbeiten. In V. Wittpahl (Ed.), *Künstliche Intelligenz* (pp. 58–71). Springer Vieweg. https://doi.org/10.1007/978-3-662-58042-4_4

Knapič, S., Malhi, A., Saluja, R., & Främling, K. (2021). Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain. *Machine Learning and Knowledge Extraction*, *3*(3), 740–770. https://doi.org/10.3390/make3030037

Koltay, T. (2017). Information Overload in a Data-Intensive World. In A. Schuster (Ed.) *Understanding Information* (pp. 197–217). Springer Cham. https://doi.org/10.1007/978-3-319-59090-5_10

Koo, Y. H., Shin, K. E., Park, J. S., Lee, J. W., Byun, S., & Lee, H. (2021). Extravalidation and reproducibility results of a commercial deep learning-based automatic detection algorithm for pulmonary nodules on chest radiographs at tertiary hospital. *Journal of Medical Imaging and Radiation Oncology*, *65*(1), 15–22. https://doi.org/10.1111/1754-9485.13105

Kozuka, T., Matsukubo, Y., Kadoba, T., Oda, T., Suzuki, A., Hyodo, T., Im, S., Kaida, H., Yagyu, Y., Tsurusaki, M., Matsuki, M., & Ishii, K. (2020). Efficiency of a computer-aided diagnosis (CAD) system with deep learning in detection of pulmonary nodules on 1-mm-thick images of computed tomography. *Japanese Journal of Radiology*, *38*(11), 1052–1061. https://doi.org/10.1007/s11604-020-01009-0

Kraus, T., Gaschow, L., Eisenträger, M., & Wischmann, S. (2021). *Erklärbare KI - Anforderungen, Anwendungsfälle und Lösungen*. iit-Institut für Innovation und Technik in der VDI/VDE Innovation + Technik GmbH. https://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/Publikation/KI-Inno/2021/Studie_Erklaerbare_KI.pdf;jsessinid=C94CE1AF540F11B6D64E3EB19A61A2AF?__blob=publicationFile&v=9

Langer, M., König, C. J., & Busch, V. (2021). Changing the means of managerial work: Effects of automated decision support systems on personnel selection tasks. *Journal of Business and Psychology*, *36*(5), 751–769. https://doi.org/10.1007/s10869-020-09711-6

Lausberg, C. & Krieger, P. (2021). Decision Support Systems in Real Estate: History, Types and Applications. In F. J. Roberts (Eds.) *Decision Support Systems: Types, Advantages and Disadvantages* (pp.1–77). Nova.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer.

Lee, K. H., Goo, J. M., Park, C. M., Lee, H. J., Jin, K. N. (2012). Computer-Aided Detection of Malignant Lung Nodules on Chest Radiographs: Effect on Observers' Performance. *Korean Journal of Radiology*, *13*(5), 564–571. https://doi.org/10.3348/kjr.2012.13.5.564

Lee, M. H., Siewiorek, D. P., Smailagic, A., Bernardino, A., & Bermúdez i Badia, S. (2021). A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–14). ACM. https://doi.org/10.1145/3411764.3445472

Li, D., Pehrson, L. M., Lauridsen, C. A., Tøttrup, L., Fraccaro, M., Elliott, D., Zając, H. D., Darkner, S., Carlsen, J. F., & Nielsen, M. B. (2021). The Added Effect of Artificial Intelligence on Physicians' Performance in Detecting Thoracic Pathologies on CT and Chest X-ray: A Systematic Review. *Diagnostics*, *11*(12), 2206. https://doi.org/10.3390/diagnostics11122206

Liao, P. H., Hsu, P. T., Chu, W., & Chu, W. C. (2015). Applying artificial intelligence technology to support decision-making in nursing: A case study in Taiwan. *Health Informatics Journal*, *21*(2), 137–148. https://doi.org/10.1177/1460458213509806

Liu, K., Li, Q., Ma, J., Zhou, Z., Sun, M., Deng, Y., Tu, W., Wang, Y., Fan, L., Xia, C., Xiao, Y., Zhang, Y., & Liu, S. (2019). Evaluating a fully automated pulmonary nodule detection approach and its impact on radiologist performance. *Radiology: Artificial Intelligence*, *1*(3), 1–9. https://doi.org/10.1148/ryai.2019180084

Martini, K., Blüthgen, C., Eberhard, M., Schönenberger, A. L. N., De Martini, I., Huber, F. A., Barth, B. K., Euler, A., & Frauenfelder, T. (2021). Impact of Vessel Suppressed-CT on Diagnostic Accuracy in Detection of Pulmonary Metastasis and Reading Time. *Academic Radiology*, *28*(7), 988–994. https://doi.org/10.1016/j.acra.2020.01.014

Mayer, A., Strich, F., & Fiedler, M. (2020). Unintended Consequences of Introducing AI Systems for Decision Making. *MIS Quarterly Executive*, *19*(4), 239–257. https://doi.org/10.17705/2msqe.00036

Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent Agent Transparency in Human–Agent Teaming for Multi-UxV Management. *Human Factors: The Journals of the Human Factors and Ergonomics Society*, *58*(3), 401–415. https://doi.org/10.1177/0018720815621206

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Mutasa, S., Sun, S., & Ha, R. (2020). Understanding artificial intelligence based radiology studies: What is overfitting? *Clinical Imaging*, *65*, 96–99. https://doi.org/10.1016/j.clinimag.2020.04.025

Nam, J. G., Kim, M., Park, J., Hwang, E. J., Lee, J. H., Hong, J. H., Goo, J. M., & Park, C. M. (2021). Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs. *European Respiratory Journal*, *57*(5). https://doi.org/10.1183/13993003.03061-2020

Nicodeme, C. (2020). Build confidence and acceptance of AI-based decision support systems - Explainable and liable AI. In *2020 13th International Conference on Human System Interaction* (pp. 20–23). IEEE. https://doi.org/10.1109/HSI49210.2020.9142668.

Oppermann, L., Boden, A., Hofmann, B., Prinz, W., & Decker, S. (2019). Beyond HCI and CSCW: Challenges and useful practices towards a human-centred vision of AI and IA. In *Proceedings of the Halfway to the Future Symposium 2019* (pp. 1–5). ACM.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*(71). https://doi.org/10.1136/bmj.n71

Papenmeier, A., Englebienne, G., & Seifert, C. (2019). holIn Proceedings of the IJCAI 2019 Workshop on Explainable Artificial Intelligence (pp. 94–100). arXiv. https://doi.org/10.48550/arXiv.1907.12652

Phillips-Wren, G., & Adya, M. (2020). Decision making under stress: The role of information overload, time pressure, complexity, and uncertainty. *Journal of Decision Systems*, *29*(1), 1–13. https://doi.org/10.1080/12460125.2020.1768680

Power, D. J., Burstein, F., & Sharda, R. (2011) Reflections on the Past and Future of Decision Support Systems: Perspective of Eleven Pioneers. In D. Schuff, D. Paradice, F. Burstein, D. Power, R. Sharda (Eds.), *Decision Support* (pp. 25–48). Springer. https://doi.org/10.1007/978-1-4419-6181-5_2

Rajpurkar, P., O’Connell, C., Schechter, A., Asnani, N., Li, J., Kiani, A., Ball, R. L., Mendelson, M., Maartens, G., van Hoving, D. J., Griesel, R., Ng, A. Y., Boyles, T. H., & Lungren, M. P. (2020). Chexaid: Deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV. *NPJ Digital Medicine*, *3*, 115. https://doi.org/10.1038/s41746-020-00322-2

Rau, R., Schweden, F., Hoppe, J. & Hacker, W. (2021). *Verfahren zur Tätigkeitsanalyse und -gestaltung bei menta-len Arbeitsanforderungen (TAG-MA)*. Asanger.

Rohmert, W. (1984). Das Belastastungs-Beanspruchungs-Konzept. *Zeitschrift Für Arbeitswissenschaft*, *38*, 193–200.

Roth, G., Schulte, A., Schmitt, F., & Brand, Y. (2020). Transparency for a Workload-Adaptive Cognitive Agent in a Manned–Unmanned Teaming Application. *IEEE Transactions on Human-Machine Systems, 50*(3), 225–233. https://doi.org/10.1109/thms.2019.2914667

Saxena, D., & Lamest, M. (2018). Information overload and coping strategies in the big data context: Evidence from the hospitality sector. *Journal of Information Science*, *44*(3), 287–297. https://doi.org/10.1177/0165551517693712

Shin, D., Zhong, B., & Biocca, F. A. (2020). Beyond user experience: What constitutes algorithmic experiences?. *International Journal of Information Management, 52*, 1–11. https://doi.org/10.1016/j.ijinfomgt.2019.102061

Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media*, *64*(4), 541–565. https://doi.org/10.1080/08838151.2020.1843357

Shrivastav, H., & Kongar, E. (2021). Information Overload in Organization: Impact on Decision Making and Influencing Strategies. In *2021 IEEE Technology & Engineering Management Conference - Europe* (pp. 1–5). IEEE. https://doi.org/10.1109/temscon-eur52034.2021.9488649

Singh, R., Kalra, M. K., Homayounieh, F., Nitiwarangkul, C., McDermott, S., Little, B. P., Lennes, I. T., Shepard, J., & Digumarthy, S. R. (2021). Artificial intelligence-based vessel suppression for detection of sub-solid nodules in lung cancer screening computed tomography. *Quantitative Imaging in Medicine and Surgery*, *11*(4), 1134–1143. https://doi.org/10.21037%2Fqims-20-630

Stowers, K., Kasdaglis, N., Rupp, M. A., Newton, O. B., Chen, J. Y. C., & Barnes, M. J. (2020). The IMPACT of Agent Transparency on Human Performance. *IEEE Transactions on Human-Machine Systems*, *50*(3), 245–253. https://doi.org/10.1109/thms.2020.2978041

Sung, J., Park, S., Lee, S. M., Bae, W., Park, B., Jung, E., Seo, J. B., & Jung, K. H. (2021). Added Value of Deep Learning–based Detection System for Multiple Major Findings on Chest Radiographs: A Randomized Crossover Study. *Radiology*, *299*(4), 450–459. https://doi.org/10.1148/radiol.2021202818

Tang, V., Siu, P. K. Y., Choy, K. L., Lam, H. Y., Ho, G. T. S., Lee, C. K. M., & Tsang, Y. P. (2019). An adaptive clinical decision support system for serving the elderly with chronic diseases in healthcare industry. *Expert Systems*, *36*(2). https://doi.org/10.1111/exsy.12369

Tortorella, G. L., Prashar, A., Saurin T. A., Fogliatto F. S., Antony, J., & Junior, G. C. (2022). Impact of Industry 4.0 adoption on workload demands in contact centers. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *32*(5), 406–418. https://doi.org/10.1002/hfm.20961

Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, technology & work*, *20*(3), 351–365. https://doi.org/10.1007/s10111-018-0481-3

Von Hehn, S., Cornelissen, N. I., and Braun, C. (2016), *Kulturwandel in Organisationen: Ein Baukasten für angewandte Psychologie im Change-Management*. Springer.

Vodrahalli, K., Daneshjou, R., Gerstenberg, T., & Zou, J. (2022). Do Humans Trust Advice More if it Comes from AI? In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 763–777). ACM. https://doi.org/10.1145/3514094.3534150

Walch, K. (2020, January 12). Is There A Difference Between Assisted Intelligence Vs. Augmented Intelligence? *Forbes.* https://www.forbes.com/sites/cognitiveworld/2020/01/12/is-there-a-difference-between-assisted-intelligence-vs-augmented-intelligence/?sh=a23756f26aba

Wamba-Taguimdje, S. L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, *26*(7), 1893–1924. https://doi.org/10.1108/BPMJ-10-2019-0411

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Wieland, R., & Hammes, M. (2014). Wuppertaler Screening Instrument Psychische Beanspruchung (WSIB) – Beanspruchungsbilanz und Kontrollerleben als Indikatoren für gesunde. *Psychologie des Alltagshandelns*, *7*(1), 30–50.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international* *conference on evaluation and assessment in software engineering* (pp. 1–10). ACM. http://dx.doi.org/10.1145/2601248.2601268

Yang, Y., Lure, F. Y., Miao, H., Zhang, Z., Jaeger, S., Liu, J., & Guo, L. (2021). Using artificial intelligence to assist radiologists in distinguishing COVID-19 from other pulmonary infections. *Journal of X-Ray Science and Technology*, *29*(1), 1–17. https://doi.org/10.3233/XST-200735

Zhang, D., Liu, X., Shao, M., Sun, Y., Lian, Q., & Zhang, H. (2021). The value of artificial intelligence and imaging diagnosis in the fight against COVID-19. *Personal and Ubiquitous Computing*, 1–10. https://doi.org/10.1007/s00779-021-01522-7

Zhang, K., Jia, X., & Chen, J. (2019). Talent management under a big data induced revolution: The double-edged sword effects of challenge stressors on creativity. *Management Decision*, *57*(8), 2010–2031. https://doi.org/10.1108/MD-06-2018-0711

Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 295–305). ACM. https://doi.org/10.1145/3351095.3372852

Zheng, N., Liu, Z., Ren, P., Ma, Y., Chen, S., Yu, S., Xue, J., Chen, B., & Wang, F. (2017). Hybrid-augmented intelligence: Collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering*, *18*(2), 153–179. https://doi.org/10.1631/FITEE.1700053

Table 1. Inclusion and exclusion criteria for the review.

		Inclusion criteria	**Exclusion criteria**
Soft factors	Population	Individuals were required to be employees	Individuals who were not employees
	Intervention	Use of an AI-based DSS when processing a work task	Use of a fully automated AI-based system; use of a conventional DSS that are not based on AI methods; use of an AI-based DSS when processing a private task
	Control	A control group that for e.g., does not use an AI-based DSS or uses an AI-based DSS that has different design characteristics	No control group included
	Outcome	Users' psychological load (e.g., workload) and related consequences (e.g., satisfaction with decision, task performance)	Users' physical load and related consequences; users' attitudes towards an AI-based DSS such as users' satisfaction with the system
Hard factors	Year of Publication	Published in 2012 or later	Published before 2012
	Language	English, German	Other languages, for e.g., Spanish, Chinese, Korean etc.
	Publication type	Journals; conference papers	Book chapters; magazine articles; reports; theses; dissertation

Note. AI: artificial intelligence; DSS: decision support system.

Table 2. Descriptive analysis of the included studies.

Authors	**Field of investigation**				**Subjects**	**Investigated main variables**						**Evaluated AI-based DSS**		**Experimental research design**				
						Independent variables				**Dependent variables**				**Location**		**Methods**		
							Characteristics of AI-based DSSs											
	Medicine	Military	Business	Tax authority		AI-based DSS in total	Correctness	Transparency	Timing of support			Real system	Simulated system	Labor	Online	Within-subject design	Between- subject design	2x2 factorial design
Gaube et al. (2022)	x				*N*= 223 physicians			x		C P	– Confidence in the decision made – Performance		x		x			x
Calisto et al. (2022)	x				*N*= 45 physicians	x				P	– Performance	x		x		x		
Jussupow et al. (2022)	x				*N*= 59 physicians		x		x	P	– Performance		x	x		x		
Vodrahalli et al. (2022)	x				*N*= 37 physicians	x				P	– Performance		x		x		x	
Li et al. (2021)	x				*N*= 119 physicians	x				P	– Performance	x		x	x	x	x	
Gaube et al. (2021)	x				*N*= 265 physicians		x			P	– Performance				x			x
Langer et al. (2021)			x		*N*= 122 lay people	x			x	S	– Satisfaction with the decision made		x	x				x

									PR SE E M P	– Perceived responsibility for the decision made – Self-efficacy – Enjoyment of a task – Monotony of a task – Performance						
Jacobs et al. (2021)	x				*N*= 220 physicians	x	x	x	C P	– Confidence in the decision made – Performance				x	x	
Knapič et al. (2021)	x				*N*= 60 lay people			x	P	– Performance	x			x		x
Lee M. H. Et al. (2021)	x				*N*=5 therapists	x			W	– Workload	x		x		x	
Stowers et al. (2020)		x			*N*= 53 lay people			x	W P	– Workload – Performance		x	x		x	
Roth et al. (2020)		x			*N*= 10 lay people			x	W P	– Workload – Performance	x	x	x		x	
Brauner et al. (2019)			x		*N*= 40 lay people		x		S PP P	– Satisfaction with the decision made – Perceived relative performance – Performance				x	x	
C. J. Cai et al. (2019)	x				*N*= 12 physicians	x			W	– Workload	x	x	x		x	
Didimo et al. (2018)				x	*N*= 32 workers at the IRV’s fiscal audit office	x			P	– Performance	x	x	x			x
Mercado et al. (2016)		x			*N*= 30 lay people			x	W P	– Workload – Performance		x	x		x	

Table 3. Overview of studies that examine the influence of the correctness of a system recommendation on users.

Authors	**Subjects**	**Task**	**AI-based DSS under examination**	**Experimental conditions**			**Effects on users …**			
				Correct condition	**Incorrect condition**	**Baseline condition: no system support**	**Objective task performance**	**Self-assessed task performance**	**Satisfaction with own task performance**	**Confidence with decision made**
Jacobs et al. (2021)	*N*= 220 physicians	Medical treatment choice	System gives a top-5 list of treatment advices (including a top-1 advice) for major depressive disorder	Top-5 list of treatment advice including a correct top-1 advice	Top 5 list of treatment adviceincluding an incorrect top-1 advice	x	incorrect < correct * incorrect < baseline * correct > baseline (n.s.)			incorrect < correct (n.s.) incorrect < baseline (n.s.) correct < baseline (n.s.)
Gaube et al. (2021)	*N*= 265 physicians	Medical diagnosis	System analyzes X-ray images of breasts and provides concrete information about abnormalities and a diagnostic advice on this basis	Correct diagnostic advice	Incorrect diagnostic advice		incorrect < correct *			incorrect < correct *
Jussupow et al. (2022)	*N*= 47 novice physicians	Medical diagnosis	System predicts pulmonary function values from a computed tomography (CT) scan for diagnosing chronic obstructive	Correct diagnostic advice	Incorrect diagnostic advice	x	incorrect < correct * incorrect < baseline * correct > baseline (n.s.)			

			pulmonary disease (COPD) and gives advice based on the analysis of the data and diagnostic advice: COPD / NO COPD						
Brauner et al. (2019)	*N*= 40 lay people	Business simulation game	System supports in the area of material disposition by providing advice for the number of supplies which should be ordered	Consistently provides correct advice	Initially, the system provides correct advice and then becomes erroneous (act like a system that breaks down)		*Round 1:* incorrect < correct * *Round 2:* incorrect < correct (n.s.)	*Round 1:* incorrect < correct * *Round 2:* incorrect < correct (n.s.)	*Round 1:* incorrect < correct * *Round 2:* incorrect < correct (n.s.)

Note. *= p< .05. n.s= p>.05. In the table, only the results of the study by Jussupow et al. (2022) are extended to participating novice physicians, but not for n=12 radiologists who were studied separately as an expert group. No significant values were observed for the latter group.

Table 4. Overview of studies that examine the influence of the level of transparency of a system recommendation on users.

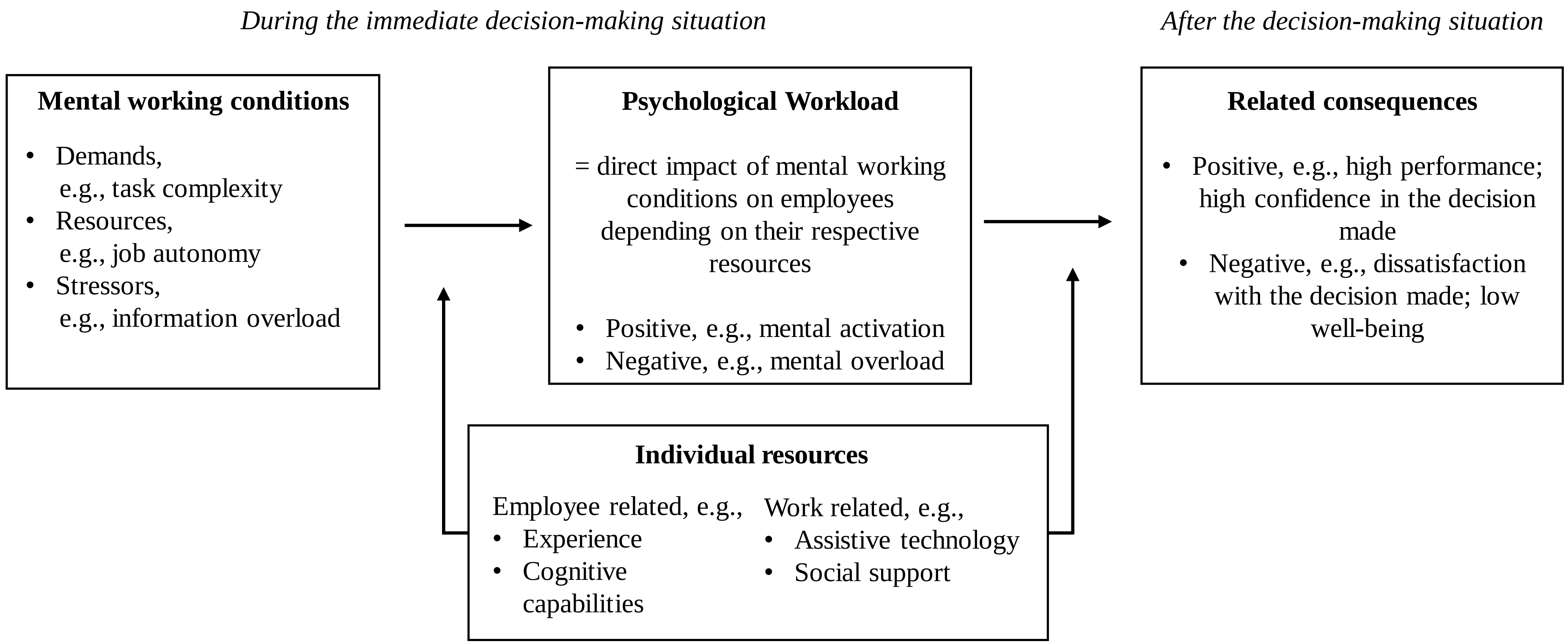
Authors	**Effect of level of transparency of an AI-based system advice on users' …**					
	Performance in terms of decision quality	**Correct AI-based advice use**	**Correct AI-based advice rejection**	**Performance in terms of time for task processing**	**Subjective Workload**	**Confidence with decision made**
Roth et al. (2020)	n.s.	n.s.	n.s.	↓ *	n.s.	
Mercado et al. (2016)	↑ *	↑ *	↑ *	n.s.	n.s.	
Stowers et al. (2020)	↑ *	↑ *	n.s.	↑ *	n.s.	
Gaube et al. (2022)	↑ *					n.s.

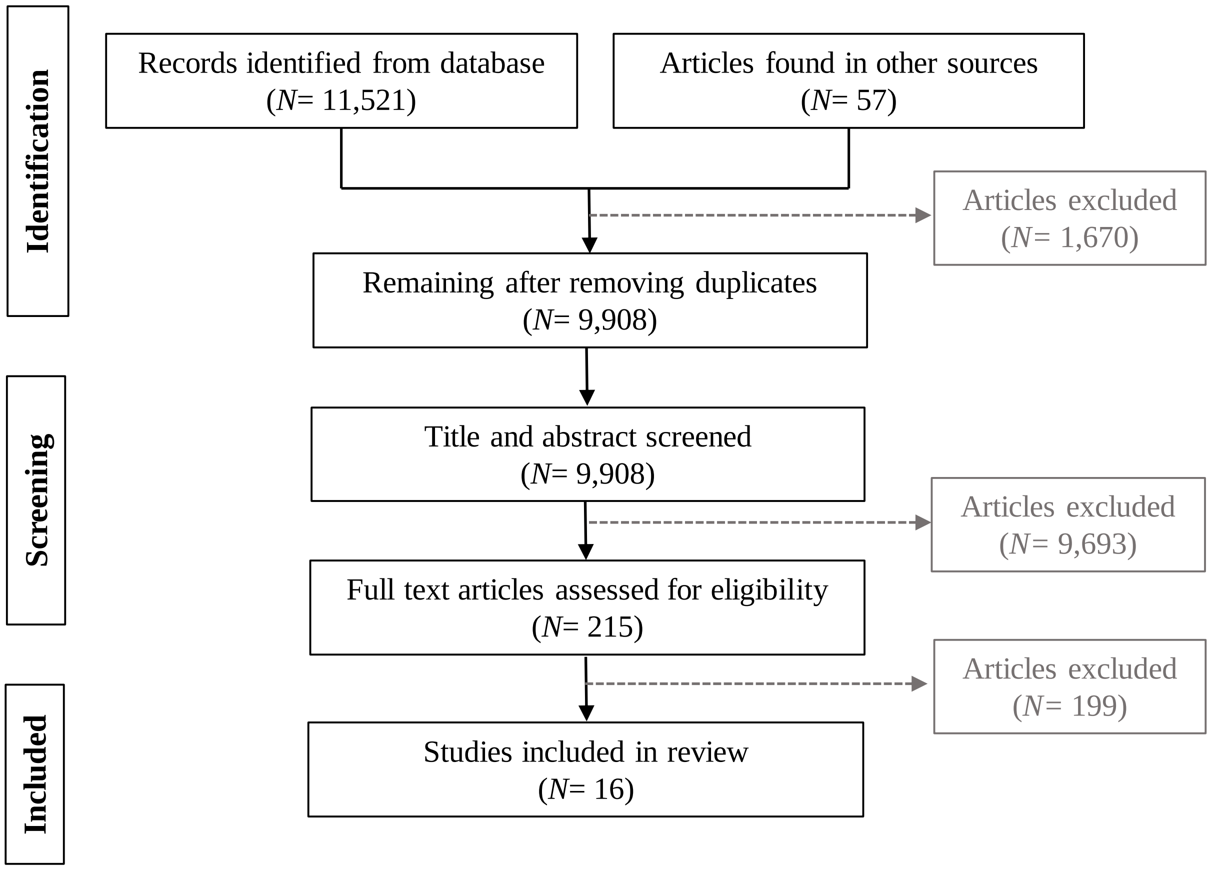
Note. ↑= increase. ↓= decrease. *= p< .05. n.s.= p>.05

Figure captions

Figure 1. Occupational psychology model of the relationship between working conditions, individual resources, psychological workload, and related consequences (adapted from Van Acker et al., 2018).

Figure 2. Flow diagram of the screening process.


During the immediate decision-making situation
After the decision-making situation
Mental working conditions
• Demands, e.g., task complexity
• Resources, e.g., job autonomy
• Stressors, e.g., information overload
Psychological Workload
= direct impact of mental working conditions on employees depending on their respective resources
• Positive, e.g., mental activation
• Negative, e.g., mental overload
Related consequences
• Positive, e.g., high performance; high confidence in the decision made
• Negative, e.g., dissatisfaction with the decision made; low well-being
Individual resources
Employee related, e.g.,
• Experience
• Cognitive capabilities
Work related, e.g.,
• Assistive technology
• Social support


Identification
Records identified from database (N= 11,521)
Articles found in other sources (N= 57)
Articles excluded (N= 1,670)
Remaining after removing duplicates (N= 9,908)
Screening
Title and abstract screened (N= 9,908)
Articles excluded (N= 9,693)
Full text articles assessed for eligibility (N= 215)
Articles excluded (N= 199)
Included
Studies included in review (N= 16)