The Soul of Work: Evaluation of Job Meaningfulness and Accountability in Human-AI Collaboration

SHADAN SADEGHIAN, University of Siegen, Germany
ALARITH UHDE, University of Siegen, Germany
MARC HASSENZAHL, University of Siegen, Germany

Work is an important part of our lives – not only as a way to earn a living but as a crucial source for experiencing meaningfulness. The introduction of autonomous systems (or in the widest sense "artificial intelligence", AI) will fundamentally impact work practices. However, while most existing models of human-AI collaboration focus on performance goals, less is known about their potential influence on job satisfaction. In this paper, we present an online experiment in which we compared the perception of job meaningfulness and accountability in a human-AI collaboration across three interaction paradigms: Supervisory, Advisory, and Interactive. Our results showed that, unlike the common notion of supervisory control, people find their job more satisfying when they directly interact with the AI and are involved and remain accountable for action and decision-making. Introducing AI as a teammate in the interactive paradigm was associated with the highest job meaningfulness.

CCS Concepts: • **Human-centered computing → Collaborative interaction**; *Empirical studies in interaction design*; *Empirical studies in collaborative and social computing*.

Additional Key Words and Phrases: Accountability; collaboration paradigm; human-automation interaction; human-AI collaboration; job satisfaction; meaningful work; automation at work; human-automation teaming.

ACM Reference Format:
Shadan Sadeghian, Alarith Uhde, and Marc Hassenzahl. 2024. The Soul of Work: Evaluation of Job Meaningfulness and Accountability in Human-AI Collaboration. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 130 (April 2024), 26 pages. https://doi.org/10.1145/3637407

1 INTRODUCTION

Work is an important part of our lives. It is not only a way to earn a living, but also an important source of meaning in everyday life [106]. The meaning of work has different sources: success on the job, the mastering of interesting challenges, good relationships with colleagues, and the feeling of being a crucial part of an organization [8]. Work has always been mediated by technology, from early hand tools to the now ubiquitous software and computers. Traditionally, these tools are passive and extend the body and the mind of the working person – the become extensions of the self. However, progress in adaptive automation, the proliferation of robots, and artificial intelligence (AI) in the widest sense already challenge this. Gradually, tools become collaborators with their own agency, unpredictability, and opacity [52]. This will redefine the human's role at work and thus impact the meaningfulness of work and job satisfaction. For example, autonomous

Authors' addresses: Shadan Sadeghian, shadan.sadeghian@uni-siegen.de, University of Siegen, Germany; Alarith Uhde, alarith.uhde@uni-siegen.de, University of Siegen, Germany; Marc Hassenzahl, marc.hassenzahl@uni-siegen.de, University of Siegen, Siegen, Germany.

ACM 2573-0142/2024/4-ART130
https://doi.org/10.1145/3637407

technology may make it difficult for workers to unambiguously attribute work successes to their own performance [73, 90].

In general, there are opposing views on how “AI” and humans should interact. While research in the AI communities often implies replacing people with autonomous systems, the Human-Computer Interaction (HCI) communities rather search for ways to augment and empower people through these systems. However, it is not yet clear what an "augmenting" and “empowering” human-AI interaction constitutes. Some researchers propose the notion of human-AI teams and see both as independent interaction partners or colleagues (also known as the *interactive* paradigm) [63, 80]. However, others believe that humans should always be in control and *supervise* AI systems [99]. In fact, most of the available interaction models for automation and human-AI interaction follow supervisory and advisory control paradigms. In these paradigms, humans monitor the actions of the autonomous system and intervene solely to rectify errors (i.e., supervisory) or validate the system’s recommendations (i.e., advisory). In these paradigms, humans are excluded from a (large) proportion of the genuine task. This leads to ambiguity, which in turn fosters disengagement and reduces feelings of responsibility [38]. Despite the fact that in both paradigms humans are excluded from the process and the reasons behind most of the immediate decisions made, they are held accountable for the outcomes of the autonomous system. This leads to so-called *responsibility gaps* [56]. Becoming more distanced from the outcomes of one’s work leads to decreased task integrity and to a lower perception of meaningfulness [9]. The problem of accountability, however, is not only limited to when the collaboration outcome is negative. One of the sources of meaning at work is achieving good outcomes and recognition of good work. Delegating tasks once performed by oneself leads to *achievement gaps*. Through extensive collaboration with autonomous systems, humans will experience fewer chances to engage in tasks that are perceived as genuine accomplishments or to receive praise [27].

In human-human collaboration, increased accountability can lead to more job satisfaction [107]. However, in such scenarios, the roles, relationships, and responsibilities of each individual are typically defined by established workplace rules and norms. It remains an open question whether these findings can be extended to human-AI collaboration. Typical questions are who should be responsible for explaining the rationale behind decisions - the AI system or the human – or who should receive praise or penalty in the cases of success or failure.

The predominant focus on performance-related aspects, such as the efficiency and effectiveness of human-automation interactions, misses the experiential aspects of working with artificial intelligence and its impact on meaningfulness. Furthermore, despite all efforts to present these systems as social entities rather than as tools [117], less is known about how rendering technology as a social counterpart impacts the relationship between human and AI, accountability, and job satisfaction. To advance our understanding of these impacts on meaningfulness and accountability, this paper aims to answer the following research questions:

- **RQ1:** How do particular paradigms of human-AI collaboration (supervisory, advisory, interactive) impact job meaningfulness?
- **RQ2:** How do particular paradigms of human-AI collaboration (supervisory, advisory, interactive) impact the perceived relationship to the AI systems and notions of accountability?
- **RQ3:** How do accountability and job meaningfulness relate to each other in human-AI collaboration?

To answer these questions, we conducted an online vignette study, in which we asked participants to assess their perception of work meaningfulness, relationship with an AI collaborator, and accountability in three different collaboration paradigms (supervisory, advisory, interactive) with different outcomes (success, failure). We found that the interactive collaboration paradigm, rather than the commonly used supervisory control, resulted in a higher perception of job meaningfulness and satisfaction. People prefer to be involved in making decisions and performing actions and render account for their outcomes. Furthermore, in a second part of the study we focused on how participants perceived their relationship to the AI. Here, the perception of AI as a teammate rather than a subordinate was associated with feelings of meaningfulness at work.

In the following, we start with a summary of related work on meaningful work, human-automation collaboration, and accountability. Then we present our experiment design followed by the results. Finally, we discuss our findings and conclude with a reflection on future work.

2 BACKGROUND AND RELATED WORK

2.1 Work as a Source of Meaning

Having experiences that give life a purpose and are perceived as meaningful makes life worth living[14, 72]. Therefore, as most people spend a large proportion of their lives at work, they do not only see it as a source of financial income but also for meaning in life, that is, workers *"search for daily meaning as well as daily bread"* [106]. Having a meaningful job that serves a higher purpose leads to higher job satisfaction [66, 69, 102] and well-being [5]. But meaningful work is not only beneficial or necessary on an individual level. It positively correlates with employees' creativity [22] and increases the likelihood that they stay at and commit to their organization [104].

Over the past five decades, understanding and designing work as a source of meaningfulness has become a topic of research[17, 47, 83, 88, 103]. There are, however, various discussions on what defines work as meaningful. Pratt and Ashforth [83] defined meaningful work as the answer to the question of *"Why am I here?"*. They believed that when an individual's work role feels as significant and is aligned with their values, work meaningfulness is fulfilled. Other definitions are *"work [that is] experienced as particularly significant"* [88, p. 95], or *"A feeling that one is receiving a return on investments in one's self in a currency of physical, cognitive or emotional energy that arises from undertaking work that is worthwhile, useful and valuable"* [58]. Overall, the subjective experience of meaningful work can come about if workers perceive it as significant and accompanied by positive meanings[88]. However, according to Steger et al. [104], the positive valance of meaningful work is more growth- and purpose-oriented rather than pleasure-oriented. Previous research in organizational psychology also supports this statement. For example, Dik and Duffy [31] argue that work that is perceived as meaningful and serves a higher purpose can be defined as a "calling" that enhances job satisfaction.

The overarching question is "what are the sources of meaningful work?". Several theoretical frameworks discuss its sources, causes, and experiences. One of the oldest and most applied frameworks is the *Job Characteristics Model* by Hackman and Oldham [49]. It defines work meaningfulness on an individual level, as a psychological state that stems from the relationship from job characteristics of skill variety, task significance, task identity, and a number of outcomes.

A more recent model by Lips-Wiersma and Morris [69] proposes four sources of meaningful work that not only stem from an individual level but also involve a more social perspective: 1) developing and becoming self, 2) unity with others, 3) expressing one's full potential, and 4) serving others. The *Four major pathways to meaningful work* framework by Rosso et al. [88] proposes similar sources: the self, others, the context, and the spiritual life, and further suggests seven mechanisms of meaningful work: self-efficacy, self-esteem, authenticity, purpose, transcendence, belongingness, and cultural/interpersonal sense-making. Finally, in other frameworks, factors such as opportunities for self-expression, helping others, and fulfillment of personal values were found to be effective in defining work as meaningful [16, 71].

2.2 Human-AI Collaboration

"Automation" describes technologies that (more or less) autonomously perform tasks that were previously done by humans [81]. However, automation does not come at once but rather in a step-wise evolving process. Consequently, human involvement is required to fulfill task goals. Ideally, the collaboration between human and automation combines the advantages of automation such as speed, adaptability, accuracy, with human capabilities such as spontaneous and creative problem solving [35, 121]. Aiming for this, several researchers proposed models that described the interaction between human and automation. One of the early examples is the model by Parasuraman et al. [81] that defines 10 levels of automation that answer the question of "who does what?" in each level. Another model by Endsley [34] defines 12 levels of automation focusing on information sharing and situational awareness of human and automation. Recently, Shneiderman criticized these models mentioning *"[...]increases in automation must come at the cost of lowering human control."* As a solution, he proposed a two-dimensional framework for human-centered AI (HCAI) with two axes for human control and automation. Its goal is to ensure human control while increasing automation (e.g., a joystick-controlled or tele-operated wheelchair, vs. a robotic one) [99]. Scholtz [94] took a different approach by describing the interaction between humans and robots based on roles that can be assigned to them (1) Supervisor: oversees the robot actions and guides it accordingly, (2) Operator: controls low-level actions, (3) Teammate: collaborate with the robot to perform the mission (4) Bystander: facilitates the relationship between the robot and environment, and (5) Mechanic: modifies abnormal behavior of the robot or fixes mechanical problems. With the advance of automation, these roles can be assigned to both human and automated systems. For example, a robot can train people for their future jobs. In this case, the models of levels of automation fall short in describing the interaction [1]. However, all of these models define the interaction between human and automation based on *supervisory control* in which the human is assigned to the monitoring of automation and occasionally intervenes to address errors or unexpected situations; or *advisory control*, in which the system only proposes decisions and a human evaluates them. Advanced automated systems, however, can also allow for *interactive control* (also known as mixed-initiative [4]), in which both human and automation can propose and evaluate decisions as a "team"[92, 118]. These paradigms allocate tasks between humans and automation, assigning assumed roles for the human, such as supervisor, advisee, and teammate. Nevertheless, it remains unclear whether the human's actual experiences and perceptions of roles, responsibilities and their relationship with automation align with the predefined roles set by these paradigms.

Apart from the definition of roles, collaboration paradigms also define how tasks should be distributed between actors. Initially, the models of human-automation collaboration were defined based on *static allocation* of functions in which tasks are assigned by either (a) choosing the better performer, (b) automating as much as possible and allocating the rest to the human, or (c) automating only when it is cost-effective [89]. However, these methods only focus on improving performance and reducing costs. Other contextual factors that influence the interaction between human and automation are disregarded. Unsurprisingly, static allocation results in diminishing humans' autonomy or assigning uninteresting tasks to them that the designers of the system determined for them [1]. To address these limitations, *adaptive allocation* was proposed and shaped the foundation of research fields such as human-autonomy (machine) teaming [25, 95, 98], and human-machine symbiosis [68]. Nevertheless, most of these approaches, still, follow the principle of allocating functions to optimize the performance in the human-automation collaboration (e.g., by adapting the system to the cognitive abilities of the human).

Recently, a few researchers investigated human-automation collaboration beyond performance-related metrics. For instance, Smids et al. [101] elicited five aspects of meaningful work with robots from a literature review, resembling the general sources for meaningful work outlined above (e.g., [47]): exercising skills and self-development, pursuing a purpose, autonomy, self-esteem and recognition, and social relationships. For each of these aspects, they discussed how introducing robots to the workspace can be a threat or opportunity for work meaningfulness. Later on, some researchers investigated the effect of each of these aspects on the perception of meaningfulness. For example, Ficuciello et al. [39] studied the effect of robot autonomy on the perception of meaningful human control in collaboration with surgical robots. Their results showed that meaningful human control should adapt to contextual factors and be framed based on the levels of robot autonomy.

Other studies explored the effect of the relationship between human and automation on work meaningfulness. Walliser et al. [113] found that introducing an AI agent as a teammate can improve affect. Breazeal et al. [12] proposed a framework based on joint intention theory [21] to ensure that the robot can anticipate its human partner's goals and needs and proactively offers help. Finally, to understand whether a robot can be a good colleague at all, Nyholm and Smids [80] compared robots in three different roles: colleague, friend, and romantic partner. They mused that a robot can fulfill many criteria for being a good colleague (e.g., working together to achieve desired goals and being reliable) but not necessarily a friend or romantic partner. In CSCW, however, most of the existing literature on human-AI collaboration studied the perception of AI as tools (e.g., [37, 53, 84]). Only a few studies that focused on the perception of AI as teammates [123] showed that when AI is perceived as a teammate, people expect it to act in different ways than a tool, for example, by showing high communication capabilities, or human-like performance.

In sum, the majority of existing research on human-automation collaboration predominantly adopts supervisory or advisory control paradigms, with a primary emphasis on enhancing performance metrics. However, the available literature exploring the meaningfulness of human-automation collaboration remains limited, often lacking empirical support. Moreover, in most of the existing studies, automated or AI-based systems are primarily portrayed as tools controlled by humans. The fewer studies that explore quasi-social relationships with these systems often overlook the underlying paradigms shaping these interactions and their influence on the perception of job meaningfulness.

2.3 Accountability in Human-AI Collaboration

Accountability is generally defined as giving or demanding reasons for one's actions to some authority [78, 85]. Bovens [11, p. 447] describes accountability as *"a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences."* Over the last decades, with the increasing application of algorithms in a variety of fields and organizations, algorithmic/AI accountability has gained a lot of attention. Based on the definition by Bovens, in a literature survey, Wieringa [119] introduced five elements that define algorithmic accountability: (1) the actor, (2) the forum, (3) the relationship between the two, (3) the content and criteria of the account, and finally (5) the consequences which may result from the account. The actor is the answer to the question of who is rendering the account. Or, who is responsible for the harm and damage that the system causes when it works correctly [74] or incorrectly [122]? Here, actors can be distinguished by three roles: decision makers (who decide about the system, its specifications, and crucial factors), developers, and users. In this paper, since we are interested in the collaboration between human and AI, we mainly address the actor as a user. The forum defines to whom the account is directed. The general notion is that once a task is delegated by a forum to an actor, the actor has to account for their conduct. In this sense, it might be reasonable to hold algorithmic systems/AI accountable. This has led to the formation of the field of explainable AI (XAI) [41], in which many researchers made efforts to provide explanations of the logic that leads to the outcome of the algorithms and to make it transparent [67]. While the transparency of the system can enhance performance or trust, it is not sufficient for holding account. Transparency is a passive construct that does not justify the reliability of a system or the rationale behind the decisions that it makes [77, p. 102], [20, p. 1177] ("see for yourself how it works"). Accountability, in contrast, demands being involved and active ("let me tell you how it works, and why") [119]. Transparency and explainability do not only impact the usability of and trust in AI-based systems. Working with "black box" AIs can decrease workers' engagement in work processes, skill cultivation, and understanding of functions and decisions of the system, and consequently reduce feelings of competence. Such systems bring users in situations where they are expected to understand and explain the actions and decisions of an AI they do not understand, but simultaneously are highly dependent on. This loss of feelings of competence, and the opaque definitions of accountability reduce the perception of meaningfulness at work [9, 26].

Collaborating with AIs that create biased, unjust, or harmful outcomes, can hold individuals accountable for actions and decisions that do not align with their morals and values and do not give them a feeling of serving others. Task significance is one of the sources of job meaningfulness [47]. Thus, the diminished perception of performing a significant task degrades overall work meaningfulness[9]. Even if the outcome of collaboration with AI is positive, people might feel a decrease in task significance due to being excluded from the action and decision process. Subsequently, workers feel distanced from those they serve, and less directly responsible, or creditable [10].

In summary, prior research on accountability in human-AI interaction has primarily relied on supervisory and advisory collaboration paradigms ((e.g., [97, 100]). These paradigms prescribe predefined roles for both humans and AI, with the aim of optimizing AI performance within controlled environments, often holding humans accountable despite their exclusion from decision-making processes. While a limited number of studies (e.g., [67, 92]) have explored accountability within interactive paradigms, their focus has been predominantly on factors like error reduction [100], performance, reaction time [67, 97], and engagement in cooperation [92]. This strong emphasis on performance metrics has given rise to what is referred to as a "responsibility gap" in human-AI collaboration, wherein the accountability for both positive and negative outcomes becomes ambiguous, along with "achievement gaps" stemming from uncertainty about who should be credited for positive outcomes. These factors are identified as key contributors to the erosion of job meaningfulness in AI collaborations [80].

3 METHOD

We conducted an experimental vignette study [2, 6, 115] to understand the effect of roles and responsibilities on the perception of work meaningfulness and accountability. Experimental vignette methodology is a widely used approach [65] that allows for controlling variables of interest which is not fully achievable in field experiments. In this method, carefully constructed and realistic scenarios are presented to participants to assess dependent variables including intentions, attitudes, and behaviors. Consequently, it enhances experimental realism and also allows researchers to manipulate and control independent variables [2, p. 2]. Our study was designed in form of a questionnaire created using SurveyMonkey [1] and distributed through Prolific [2]. The questionnaire was in English.

Previous CSCW research [123] shows that to gain an in-depth comprehension of human-AI collaboration, it is essential to explore the interaction in a concrete domain with a specific context.

[1]https://www.surveymonkey.com
[2]https://www.prolific.com/

Therefore, we asked participants to imagine scheduling the work shifts for nurses at a hospital. We selected shift scheduling as a scenario for a couple of reasons. First, it plays a significant role for stakeholders, including management, healthcare workers, and care recipients. Thus, the decisions have important consequences for planners and their relationships with other workers. Second, finding suitable schedules involves carefully balancing economic, legal, and social interests. This complexity makes manual scheduling time-consuming and demanding for planners, but also implies that there is no “one best” solution. A shift plan optimized for efficiency is different from one optimized for workers’ well-being [108], and these priorities are set by humans. In other words, a mix of (explicitly) human-made and automated (partial) decisions seems appropriate for this task [109, 110]. And third, while automatic shift scheduling has been extensively researched for decades, its practical integration in real-world settings remains a challenge [59, 82], in part because fully automated, abstracted approaches tend to overlook context-specific and social affordances [45, 96]. From the perspective of healthcare workers, shift scheduling impacts both their professional and personal lives. Irregular or long shifts have negative health effects [3], while specific shift assignments (e.g., on weekends or holidays) can disrupt social life [23]. The success of a shift often relies on considering individual factors such as fatigue, personal plans, and relationships with co-workers. Thus, although the schedules created by algorithms might be optimal regarding the performance metrics, they might lead to undesired outcomes due to neglecting the preferences, thoughts, and feelings of the nurses [109, 110]. The currently existing shift-scheduling AI-based systems, which identify counterintuitive patterns and prompt humans to make adjustments, follow a supervisory control approach. In this approach, users are involved in error correction without having access to the algorithm’s decision-making parameters. Consequently, individuals may perceive a reduced sense of accountability for the overall outcome, which can impact their perception of meaning in the context of their work. Therefore, we chose shift scheduling as a test scenario to explore the interplay between human involvement, and the resulting perceptions of meaning and accountability. In our study, the shift scheduling task consisted of six sub-tasks. We defined these sub-tasks in line with [28, 108, 110].

Our study had a two-factorial mixed design, with the between-subjects factor “outcome” (Success, Failure) and the within-subjects factor “Collaboration paradigm” (Supervisory, Advisory, Interactive). Half of the participants read through a version of the scenario in which collaboration with AI led to success, and for the other half, it led to failure. Furthermore, each participant went through three different collaboration paradigms (CP):

- *Supervisory:* human is primarily dedicated to monitoring the AI and intervenes occasionally to address errors or unexpected circumstances.
- *Advisory:* AI makes suggestions about possible actions, no action will be undertaken without human confirmation, and
- *Interactive:* both human and AI can make decisions and perform actions.

Table1 describes the vignettes for the collaboration paradigm and outcome conditions that we showed to the participants. We provided a fine-grained definition of collaboration paradigms to ensure that participants recognize the differences between different paradigms (levels of the variable), and enhance the immersiveness of the vignettes by providing more details as suggested by Aguinis and Bradley [2]. However, we gave an abstract definition of the sub-tasks within each paradigm to make them understandable to participants that are not familiar with shift planning for nurses. The vignettes were presented in a counterbalanced order. After reading each vignette, the participants filled in a questionnaire (described below).

<table>
<tr><td colspan="3">The head of your department has assigned you to schedule the working shifts of nurses on a weekly basis. Around 70 nurses with different labor hours per week and personal preferences work in your department. You will be doing this task together with an artificial-intelligence powered system called “CareShifts” the hospital has recently purchased. CareShifts is specialized in planning shifts based on work regulations, activity demands, and employees’ needs. It is highly reliable and provides high-quality solutions</td></tr>
<tr>
<td>Supervisory:
CareShifts
• Defines work standards: how many patients a nurse can check in a shift
• Forecasts future activity levels: it identifies peak and slow times during each day, and each week of the month
• Calculates the exact number of nurses needed per shift
• Considers scheduling regulations: e.g., maximum number of working hours per week for a nurse, minimum number of nurses per shift, minimum break hours between two shifts
• Considers employee scheduling needs: desire for flexibility, desired rotation times
Creates the schedule
You
• Monitor the performance of CareShifts
• Intervene and overwrite CareShifts’ actions when you detect errors or risk of undesired outcomes</td>
<td>Advisory:
CareShifts
• Defines work standards: how many patients a nurse can check in a shift
• Forecasts future activity levels: it identifies peak and slow times during each day, and each week of the month
• Calculates the exact number of nurses needed per shift
• Considers scheduling regulations: e.g., maximum number of working hours per week for a nurse, minimum number of nurses per shift, minimum break hours between two shifts
• Considers employee scheduling needs: desire for flexibility, desired rotation times
• Creates the schedule
In all steps CareShifts proposes solutions but waits for your confirmation to proceed to the next step.
You
• Check the solutions proposed by CareShifts in each step
• Confirm, or reject and overwrite the proposed solution</td>
<td>Interactive:
• You define work standards: how many patients a nurse can check in a shift
• CareShifts forecasts future activity levels: it identifies peak and slow times during each day, and each week of the month
• You calculate the exact number of nurses needed per shift
• CareShifts considers scheduling regulations: e.g., maximum number of working hours per week for a nurse, minimum number of nurses per shift, minimum break hours between two shifts
• You consider employee scheduling needs: desire for flexibility, desired rotation
• You and CareShifts create the schedule together.</td>
</tr>
<tr><td colspan="3">You and CareShift completed the scheduling for this week and shared it with the nurses. An hour later, you received 10 praise (success)/complaint(failure) emails from the nurses, the head of the department in cc. The nurses voice their satisfaction (success)/dissatisfaction(failure) with the shift schedule.</td></tr>
</table>

Table 1. Description of vignettes for collaboration with AI for shift planning with supervisory, advisory, and interactive collaboration paradigms, and success and failure outcome conditions

3.1 Measures

3.1.1 *Job Meaningfulness and Affect.*

Our questionnaire had two parts. The questions in the first part were regarding job meaningfulness and satisfaction. For this part, after reading the vignettes in each condition, we asked the participants to fill in the Job Diagnostic Survey (JDS) (Cronbach's α= .70) by [47, 48] which is a tool for measuring experienced meaningfulness at work. The JDS consists of five scales; *skill variety (SV)* (5 items, Cronbach's α= .73): the degree to which the job involves a variety of activities and requires the worker to learn different skills, *task identity (TI)* (4 items, Cronbach's α= .60): the degree to which the job entails completing an identifiable task with a clear outcome, *task significance (TS)* (4 items, Cronbach's α= .64): the degree to which the job impacts the lives and works of others, *autonomy (AU)* (4 items, Cronbach's α= .66): the degree to which the job provides the worker freedom, independence, and autonomy to make decisions and perform actions, and *feedback (FB)* (6 items, Cronbach's α= .66): the degree to which workers receive clear and applicable information about their performance at a job. All items were rated on a 5-point Likert scale (1= Very non-descriptive, 5 = Very descriptive). Furthermore, in the JDS we can calculate an overall Motivating Potential Score by using this formula:($MPS = \frac{(SV+TI+TS)}{3} \times AU \times FB$) [47]. All scales have an acceptable level of reliability [111]. Therefore, we included all of them in our analysis.

Since JDS does not directly ask about the perception of job meaningfulness and satisfaction, which was the core subject of our study, we asked participants to rate their perception of meaningfulness ("How meaningful do you find your task?"), and satisfaction ("How satisfied are you with your task?") in their job. Furthermore, to understand how the collaboration paradigm and outcome change the perception of the performance of the AI, we asked the question ("How well-performing do you perceive CareShifts?"). Participants answered these three questions on a 7-point Likert scale. Moreover, for each vignette, we asked participants what they found positive/negative about their job (open question) and how positive/negative they find working with CareShifts on a 7-point Likert scale.

3.1.2 *Accountability and Relationship to AI.*

The second part of the questionnaire was dedicated to the perception of accountability and the relationship between human and AI. Based on Boven's definition of accountability [11], we asked participants in three 9-semantic differential scale questions (1=You, 5=Both, 9=CareShifts) *"Who is accountable for the shift schedule?"*, *"Who should be explaining the rationale behind the decisions for making the schedule to the head of the department or nurses?"*, and *"Who should be [praised/penalized] for the outcome?"*.

As mentioned above, previous research has shown that accountability is a construct defined by human involvement in the interaction or the responsibilities assigned to each interaction partner. These allocated responsibilities also influence the relationship between the actors[119]. Therefore, to understand how they perceived their relationship with CareShifts, participants were asked to rate three statements mentioning *"In this scenario, CareShifts feels as a [superior/subordinate/teammate] to me"* on a 7-point Likert scale (1= not at all, 7= extremely). Finally, for each vignette, we asked the participants how well they could imagine the presented situation on a 7-point Likert scale (1= not at all, 7= extremely).

3.2 Participants

51 participants (28 female, 4 diverse) aged between 18-74 years (m=30.6, sd= 13.2), recruited through Prolific, filled in the questionnaire completely. Their occupations were knowledge workers (17), student (14), self-employed (6), teacher (4), IT professional (4), unemployed (2), retired (2), therapist (1), and undefined (5). We set English language fluency as a pre-screening criterion. Prolific required

Scale	Outcome	Supervisory		Advisory		Interactive	
		Mean	SD	Mean	SD	Mean	SD
Motivating Potential Score	Success	42.42	18.95	46.43	21.04	59.73	22.27
	Failure	39.14	18.07	42.37	18.80	48.20	23.72
Skill Variety	Success	2.96	0.87	2.88	0.87	3.78	0.69
	Failure	2.93	0.80	3.10	0.66	3.29	0.72
Task Significance	Success	4.25	0.40	4.36	0.70	4.62	0.52
	Failure	4.10	0.73	3.90	0.94	4.05	0.78
Task Identity	Success	3.02	0.54	2.94	0.60	3.30	0.54
	Failure	3.12	0.74	3.17	0.75	3.26	0.65
Autonomy	Success	3.37	0.75	3.58	0.89	3.88	0.81
	Failure	3.35	0.76	3.51	0.70	3.67	0.79
Feedback	Success	3.48	0.72	3.67	0.60	3.85	0.60
	Failure	3.32	0.69	3.41	0.69	3.48	0.65

Table 2. Means and standard deviations of JDS ratings.

both us and the participants to agree to their terms of use which included ethical aspects for data handling. Furthermore, prior to the study, participants were informed that the survey data will not contain information that personally identifies them and will be used for scholarly purposes only. Participants were compensated with 8€/hour.

4 RESULTS

Since the paradigms tested in this experiment do not currently exist in real-life scenarios, the only way to gauge its realism was by relying on participants' subjective responses regarding how immersive it is based on their current knowledge and experience. In all conditions, participants found the scenarios immersive with a high average of ratings (M=5, SD= 1.2) on a 1-7 scale. This means that participants could relate to the vignettes as a realistic situation. In the following, we present our results according to our research questions.

4.1 RQ1: How do particular paradigms of human-AI collaboration impact job meaningfulness?

4.1.1 Job Diagnostic Survey (JDS).

Table 2 indicates the descriptive statistics for the overall score of the JDS (MPS) and each scale. Results from a 2×3 ANOVA with CP (Supervisory, Advisory, Interactive) as within-subject factor, Outcome (Success, Failure) as between-subject factor, and MPS as measure revealed a significant effect of CP, $F(2,98)=10.45$, $p<0.001$, $\eta^2 = 0.06$. A Bonferroni pairwise comparison revealed that in the Interactive condition, participants rated their job as more motivating and satisfying than in the Advisory ($t(49)= 3.26$ $p=0.006$), and Supervisory $t(49)= 3.76$ $p=0.001$) conditions. No significant main effect of the Outcome nor an interaction effect between the two independent variables were observed (Figure 1).

By conducting five similar 2×3 ANOVAs, we analyzed the differences in each scale of JDS. Similar to the results from overall MPS, we found a significant main effect of the CP on ratings for Task Identity ($F(2,98)=3.71$, $p=0.02$, $\eta^2=0.026$), Autonomy ($F(2,98)=4.90$, $p=0.009$, $\eta^2=0.044$), and Feedback

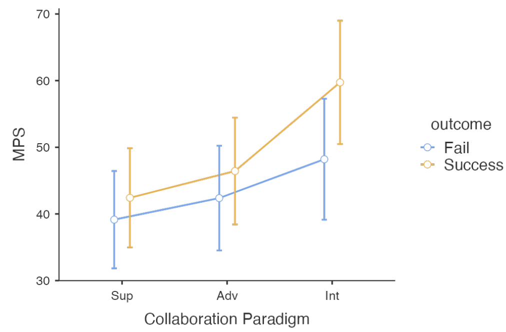

Fig. 1. Means and 95% confidence intervals of the job diagnostic survey overall motivating potential scores (MPS) across Collaboration Paradigm conditions

(F(2,98)=6.05, p=0.003, η^2=0.026). Post-hoc comparisons showed that participants believe that their job provides more autonomy and feedback in the Interactive than in the Supervisory condition $t_{AU}(49) = 2.64, p_{bonferroni} = 0.03, t_{FB}(49) = 3.25, p_{bonferroni}= 0.006$. For the Task Identity sub-scale, however, this difference was observed between the Interactive and the Advisory condition $t_{TI}(49) = 2.54, p_{bonferroni}= 0.04$. This can be explained by the higher interdependence between the human and AI tasks in the advisory condition, which leads to a lack of identifiability of each individual task.

We also observed a significant main effect of CP on ratings for the Skill Variety sub-scale (F(2,98)=17.65, p<0.001, η^2=0.10). The Interactive paradigm was rated to require significantly more skills than the Supervisory t(49)= 4.40, $p_{bonferroni}$<0.001, and Advisory t(49)= 5.04, $p_{bonferroni}$<0.001 paradigms. The significant interaction effect between CP and Outcome shows that this effect varies depending on the success of the interaction (F(2,98)= 5.34, p=0.006, η^2=0.03). When collaboration led to failure, no significant difference was found between different paradigms. In the case of a successful outcome, however, the Interactive condition was rated to need more skills than the Supervisory ($p_{bonferroni}$= 0.001) and Advisory ($p_{bonferroni}$<0.001) conditions.

Finally, we found that participants perceive the importance of their task depending on its success. A significant main effect of Outcome on Task Significance ratings (F(1,49)=5.53, p=0.023, η^2=0.07) showed that when collaboration leads to success, people perceive it as more important and contributing than when it leads to failure.

4.1.2 Job Meaningfulness and Satisfaction.

We asked participants how meaningful they found their job in each scenario. The results of a two-way ANOVA between CP and Outcome showed a significant main effect of CP on job meaningfulness (F(2,98)= 6.61, p=0.002, η^2=0.031). Participants found their job in the Interactive condition (M=5.77 SD= 0.21) significantly more meaningful than in the Supervisory condition (M= 5.03, SD= 0.21), t(49)=3.60, $p_{bonferroni}$=0.002.

Furthermore, participants were asked how satisfying they found their job in each scenario. A similar 2×3 ANOVA showed significant main effects of CP (F(2,98)= 6.15, p=0.003, η^2=0.04), and Outcome (F(1,49)= 6.07, p=0.01, η^2=0.07) on job satisfaction. In line with the meaningfulness ratings, participants perceived the Interactive condition as significantly more satisfying than the Supervisory condition (t(49)=3.32, $p_{bonferroni}$=0.005). As expected the condition with a successful outcome (M=5.75, SD=0.25) led to higher satisfaction than failure (M=4.88, SD= 0.24).

4.1.3 Perception of AI Performance.

In each condition, participants were asked how well-performing they find CareShifts. A 2×3 ANOVA showed a significant main effect of Outcome (F(1,49)= 13.5, p<0.001, η^2=0.144). As awaited, participants rated the performance of the CareShifts in success (M= 5.81, SD= 1.21) significantly higher than in failure (M=4.62, SD= 1.61) condition. No significant main effect of CP was observed. However, a significant interaction effect between Outcome and CP (F(2,98)= 3.81, p=0.02, η^2=0.024)) shows that when the collaboration led to success, participants rated the performance of CareShifts in the Supervisory paradigm better than Advisory and Interactive paradigms, respectively. This order, however, was reversed when the collaboration led to failure. Post hoc pairwise comparisons showed a significant difference between Success and Failure conditions in the Supervisory paradigm. t(49)=4.19, $p_{bonferroni}$=0.002 (Figure 2). This is quite evident, as the contribution of CareShifts in the collaboration is the highest in the Supervisory and the lowest in the Interactive paradigm. Consequently, its evaluation of performance follows this order depending on the outcome.

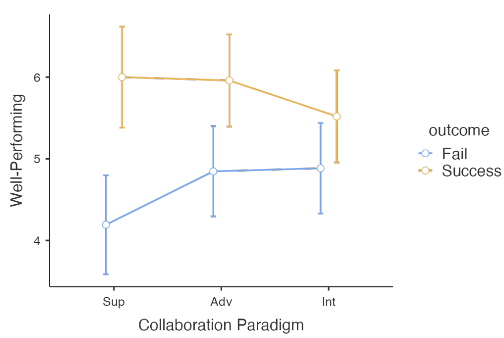

Fig. 2. Means and 95% confidence intervals of ratings for how well-performing the AI is perceived across Collaboration Paradigm and Outcome conditions

4.1.4 Positive and Negative Affect.

We asked the participants to rate how positive and negative they found their collaboration with CareShifts. On the positive scale, the results from a 2×3 ANOVA showed no significant main or interaction effect of CP and the Outcome. On the negative scale, however, we observed a significant main effect of CP (F(2,92)= 4.27, p=0.01, η^2=0.025). Post-hoc paired t-tests showed that participants perceive their collaboration with CareShifts in the Interactive condition significantly less negative than in the Supervisory condition (t(46)= 3.01, $p_{bonferroni}$<0.012). These ratings can be elaborated by the results of the open question which asked participants what they find negative about their job. In the Supervisory condition, one of the factors that made the job unappealing was the significance of the job (n=12). People did not have the impression that their job was important since it was mainly performed by CareShifts: *"I don't really have a very important job."; "The program worked out the shift based on future activity levels etc. and not me, so, therefore, I feel a bit useless"*. Other reasons were low workload and lack of challenges (N=8), and lack of stimulation (n=7) due to the design of the job, or working together with an AI. Finally, 5 participants mentioned lack of freedom and autonomy in the supervisory condition as a negative aspect.

4.2 RQ2: How do particular paradigms of human-AI collaboration impact the perceived relationship to the AI systems and notions of accountability?

In each condition, we asked participants who is accountable for the schedule (1=you, 5=both, 9=CareShifts). Here, our 2 ×3 ANOVA did not reveal any significant main effect of CP or Outcome, nor an interaction effect between them. In another similar scale, participants were asked who should be explaining the rationale behind the decisions for making the schedules to the head of the hospital or nurses. The results from a two-way ANOVA showed that there was a significant main effect of CP on the ratings (F(2,98)= 6.31, p=0.003, η^2=0.04). Post-hoc tests revealed a significant difference between the Supervisory (M=4.05, SD=2.49) and Interactive (M= 2.90, SD=2,89) conditions (t(49)=3.03, $p_{bonferroni}$=0.012). While in the Supervisory paradigm, the ratings were closer to the belief that both the human and CareShifts should be explaining the decisions, in the Interactive paradigm, the ratings were more towards the lower values implying the human should be the explaining actor.

Finally, we asked the participants who should be praised (condition Success) or penalized (condition Failure) for the outcome. A two-way ANOVA showed a significant main effect of CP (F(2,98)= 4.36, p=0.015, η^2=0.03). Similar to the first question, in the post-hoc tests, we found a significant difference between Supervisory and Interactive paradigms (t(49)= 2.66, $p_{bonferroni}$=0.03). In the Supervisory paradigm, the ratings were more in the middle of the scale implying that both human and CareShifts should be praised/penalized (M= 4.51, SD= 2). But in the Interactive paradigm, they were more towards the lower values indicating the human should be praised/penalized (M= 3.56, SD= 3.56). These results show that the participants generally believe that explaining the decisions and getting praise or penalty should be shared between both human and AI. Nevertheless, with the increase in the contribution of the human collaborator, their share in receiving the praise or penalty enhances.

To understand how the CP and Outcome influence the perceived relationship between the human and AI collaborator, we asked participants to rate their perceived relationship to CareShifts. We conducted three 2×3 ANOVAs with CP as a within-subject factor, Outcome as a between-subject factor, and relationship (superior, teammate, subordinate) as a dependent variable. The results showed a marginal effect of the CP on the perception of CareShifts as teammate (F(2,98)= 2.99, p=0.05, η^2=0.013). Participants perceived CareShifts as a teammate more in Interactive (M= 4.82, SD=2,17) than Supervisory (M= 4.31, SD= 2.19), and Advisory (M=4.25, SD= 2,32) conditions.

Our Collaboration Paradigms define relationships (e.g., supervisor) through the distribution of sub-tasks. We were interested to know whether the perception of the relationship to CareShifts really changes in each CP condition. Therefore, we conducted an exploratory 2×3×3 ANOVA with Outcome as a between-subject factor, relationship (superior, teammate, subordinate), and CP (Supervisory, Advisory, and Interactive) as within-subject factors, and the Intensity of Relationship as a dependent variable. We observed a significant main effect of Relationship $F_{Rel}(2, 98) = 22.85, p < .001, \eta^2 = 0.194$. Participants generally perceive CareShifts significantly more as a teammate (t(49)=7.07, $p_{bonferroni}$<.001), or subordinate (t(49)= 5.41, $p_{bonferroni}$<0.001) than a superior to them. No significant main effect of CP was observed (Figure 3). Nevertheless, the significant interaction effect between CP and Relationship ($F(4, 196) = 2.50, p < .04, \eta^2 = 0.01$) showed that the mean difference between teammate and superior relationships in the Interactive paradigm is larger than in the Supervisory and Advisory paradigms. Furthermore, in the Interactive condition, we also observed a larger (non-significant) mean difference between teammate and subordinate relationships than in other CPs (Table 3). No significant main effect of Outcome or interaction effect between the variables were found.

CP	Relationship	Mean	SD
Sup	Superior	2.36	0.248
	Subordinate	3.83	0.279
	Teammate	4.31	0.295
ADV	Superior	2.06	0.238
	Subordinate	4.22	0.310
	Teammate	4.25	0.326
Int	Superior	1.94	0.214
	Subordinate	3.67	0.329
	Teammate	4.82	0.305

Table 3. Means and standard deviation values for relationship intensity across Collaboration Paradigms and Relationships

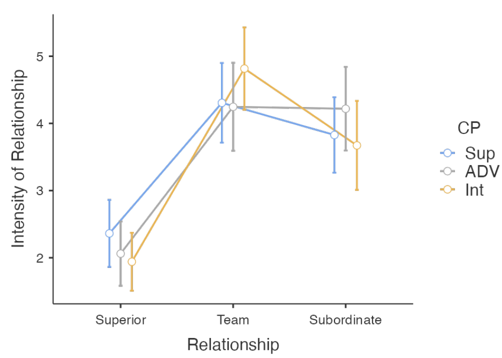

Fig. 3. Means and 95% confidence intervals of Intensity of Relationship across Collaboration Paradigm, and Relationship conditions

4.3 RQ3: How do accountability and job meaningfulness relate to each other in human-AI collaboration?

To gain a better understanding of the relationship between job meaningfulness and accountability, we conducted a set of correlation analyses between the corresponding variables. We found a significant positive correlation between the perception of AI as a Teammate and job meaningfulness (Pearson's r= 0.29, p<.001). The perception of AI as Superior, however, was negatively correlated with job meaningfulness (Pearson's r_{Sup}= -0.27, p<.001). No significant correlation was observed between the perception of AI as a subordinate and job meaningfulness.

In our questionnaire, on the accountability and explaining scales, lower values are associated with higher human accountability and contribution in explaining the decisions (1=You, 5=Both, 9=CareShifts). To increase readability, we reversed these scales for this analysis. We observed a significant positive correlation between the perception of accountability and job meaningfulness (Pearson's r= 0.28, p<.001). This means that higher human accountability corresponds with a higher perception of meaning. A similar relationship was also observed between meaningfulness and explaining the decisions ratings (Pearson's r= 0.17, p<.034). Higher account for the human to explain the decisions corresponds with higher perception of job meaningfulness.

5 DISCUSSION

In this paper, we examined how collaboration paradigms and outcomes impact job meaningfulness and accountability in human-AI collaboration using the experimental vignette methodology. In the following sections, we discuss our results.

5.1 Summary of Results

Our results showed that the collaboration paradigm of human-AI collaboration affects the perception of job meaningfulness and satisfaction. In the Interactive paradigm, participants found their job more meaningful and satisfying than in the Advisory and Supervisory paradigms. The results of the job diagnostic survey confirm and further clarify this relationship. They show that in the Interactive paradigm, the job provided more autonomy in decision-making, required more skills, was more identifiable, and allowed more feedback from the collaborator. These results imply that when people contribute more directly to the tasks and are in dialogue with their AI collaborator, they find their jobs more fulfilling and appealing than when they supervise the AI or confirm its actions. Furthermore, the participants associated their job in the Supervisory paradigm with more negative feelings due to lack of autonomy, feeling incompetent, and having the impression of doing an unimportant task. All in all, these findings are consistent with the literature on job meaningfulness [47] and extend them to the domain of human-AI collaboration.

The outcome of the collaboration also influenced the perception of job meaningfulness and the performance of AI. When the collaboration with AI was successful, participants found their job more satisfying and important. Moreover, they found AI as a better-performing collaboration partner. Given that in all our three scenarios, the AI performed a significant part of the job, it is foreseeable that its performance is evaluated by the joint outcome. The interaction effect between CP and Outcome variables on the performance ratings, however, shows that the perception of AI's performance highly depends on the extent of its contribution to the task. In conditions with a successful outcome, people perceived the AI as better-performing in Supervisory CP than in Interactive CP. Nonetheless, when the collaboration failed, the AI in Supervisory CP (with higher contribution) was rated as worse-performing than in Interactive CP. These results are observed although the performance of CareShifts was described as *"highly reliable and provides high-quality solutions"* across all conditions. An unsuccessful outcome violates participants' expectations of a "highly reliable" system. Nevertheless, the contribution of the AI to the task was an important factor in defining its share in the success or failure of the outcome.

To understand the effect of CP on accountability, based on Boven's definition of accountability [11], we asked the participants three questions about to whom the account is assigned, who renders the account, and who faces the consequences. For the first question, we did not find a significant effect of CP on accountability. In all paradigms, participants' ratings were centered around the middle point of the scale indicating that they believed both the human and AI should be held accountable regardless of the outcomes. Nevertheless, when we asked them who should explain the rationale behind the decisions (rendering the account), the results differed. Participants generally believed that both human and AI should be explaining their decisions. However, in the Interactive paradigm where the human was more involved in performing the task, they expressed that the human should be the explaining partner. One interpretation of the difference between the results of these two questions is that the first question does not convey the act of "rendering account" as strongly as the second, which emphasizes providing an explanation for the actions. Finally, we found that participants thought that when the human interacts with AI in the Supervisory paradigm, they both should be praised or penalized for the outcome, but in the Interactive paradigm, the human receives the praise or the penalty. These ratings can be also interpreted through the lens of the so-called *"Superpowers of Robots"* [116]; it is meaningless for an AI [robot] to feel proud or embarrassed. Therefore, as soon as the human's presence is more salient in the collaboration, people grant praise or penalty to him/her.

CareShifts was generally perceived more as a teammate or subordinate rather than a superior. Nevertheless, participants perceived CareShifts more as a teammate in the Interactive paradigm than in other paradigms. The reason could be the equal and interdependent distribution of tasks between human and AI. Among the three questioned relationships, only having AI as a teammate was associated with higher job satisfaction in collaboration with it. Furthermore, having a higher share in holding account for and explaining the outcomes was associated with job meaningfulness. Based on our results, which showed that the Interactive paradigm involves the human as the explaining partner to the greatest extent, these findings suggest that aiming for the Interactive paradigm could lead to higher levels of job meaningfulness.

5.2 Job Meaningfulness: From Human-Human to Human-AI Teams

Based on our results, we deduce that designing human-AI collaboration in ways that enhance the perception of the AI as a teammate leads to higher job meaningfulness. It is important to note that before AI takes a role equal to human, it should reach a level of maturity to be able to communicate and collaborate with humans [1]. When comparing collaboration with humans to AI, previous studies showed that people generally prefer working with humans rather than with AIs [42, 91]. However, in working with AI, the distribution of tasks affects the perception of meaningfulness [91]. Previous research on human-human teams defines team meaningfulness as a collective construct that is associated with meaningfulness at the individual level and the perception of one's task in the team as worthwhile [18]. Meaningfulness in human teams is related to three core job characteristics: Skill Variety, Task Identity, and Task Significance [60]. In our study, we observed that in collaboration with AI, the perception of these three characteristics and meaningfulness was higher in the Interactive paradigm, which also had the highest perception of AI as a teammate.

Besides job characteristics, other aspects of work influence the perception of meaningfulness in human-human teams. One of these aspects is the social context of work which refers to the interpersonal relationships that are defined by the roles and tasks of members of the team and play an essential part in shaping their experiences [13, 43, 70]. Studies have shown that employees perceive higher meaningfulness in jobs that enhance interaction between team members [24]. The human-human relationships formed in the workplace play a crucial role in evaluating work and future career prospects. However, as collaboration with AI becomes increasingly inevitable, there is concern about whether such collaborations can maintain these important experiences. Our results indicate that higher levels of human-AI interaction in the Interactive paradigm lead to higher ratings of job meaningfulness. Nonetheless, an important question remains regarding the extent to which collaboration with AI, which may come at the expense of human collaboration, is beneficial. The feeling of relatedness is a fundamental psychological need for humans, the fulfillment of which can shape positive and meaningful experiences [30]. Thus, collaboration with AI that sacrifices human relationships can result in negative experiences. Previous studies have shown that AI (robots) can fulfill the criteria of a colleague [80], but uncertainty remains about how the interaction should be designed to fulfill the need for relatedness. Therefore, further investigation is needed to understand how the introduction of AI teammates influences prosocial practices in the workspace, as well as feelings of relatedness and accountability for outcomes.

Finally, similar to human-human teams, previous research suggests that to classify a group of human and AI agents as a "team", an element of interdependence and identity is required [79]. This helps members gain more information about their task as a whole and enhances their task identity and their perception of being important to the team. In our study, we defined this interdependency in the vignettes (especially the Interactive paradigm) so that the outputs of the humans' tasks were the inputs for the AI's tasks and vice versa. However, we did not explore the details of the interaction between human and AI, e.g., how the human hands over the output of his/her sub-task to AI, or what is the content and presentation form of the feedback that they give/receive. Further studies are needed to explore these factors and their effect on the perception of AI as a teammate and job meaningfulness.

5.3 Paradigms for Human-AI Collaboration: Are We Really Human-centered?

The design of human-automation (AI) interaction, is often defined with the objective of performance i.e., effectiveness and efficiency in fulfilling task goals. Nevertheless, seeing the human and AI as a team, this objective can be defined as the "goal" that the team is shaped for [64]. The goal itself, however, is not the glue that holds the team together. Gross and Martin [44] defined team cohesiveness in two underlying dimensions: task cohesiveness which is the commitment or attraction to the team task [46], and interpersonal cohesiveness which is the team members' attraction to or liking of the team [36]. The models of human-automation interaction, yet, often ignore this facet and stick to notions of static allocation of tasks that mostly leave humans in the supervisory control of the automated system. This consequently, leads to assigning uninteresting tasks to humans or removing their autonomy by forcing them to do the tasks that the designer of the system planned for them [1]. Even if the distribution of tasks is more considerate (e.g., adaptive allocation), it mostly follows the principle of allocating functions to optimize performance by, for example, designing automation in accordance with the cognitive abilities of humans (e.g., [81, 99]). Interestingly, while this form of interaction is proposed to ensure reliability, previous research has shown that humans are prone to errors in monitoring and vigilance due to reasons such as boredom, fatigue, distraction, lack of situational awareness, and deskilling [51, 114]. Furthermore, supervisory control paradigms, often define the automated system as a "tool" (or subordinate to humans) [99]. However, with the increase in complexity, opaqueness, and agency, automated systems can appear to humans as counterparts rather than tools [52]. Ignoring these quasi-social collaboration paradigms can lead to missing out on factors that influence the interpersonal (in our case inter-member) cohesiveness in human-automation teams.

Our results showed that the outcome of the human-AI collaboration has a significant effect on job satisfaction and perception of the performance of the AI. Thus, fulfilling a performance-driven goal can contribute to human well-being. Moreover, previous research has shown that perceiving AI as a well-performing and reliable collaborator, engenders increased trust in the AI [54]. In our study, however, we found that while performance is beneficial to humans' perception of job meaningfulness, it is not the only influential factor. The paradigm that defines the collaboration between human and AI, has a strong effect on how meaningful the job is perceived. Our results implied that the widely-applied supervisory control paradigm leads to lower job satisfaction and meaningfulness than the interactive paradigm. One reason could be that shifting the role of humans from decision-makers and actors to supervisors drastically changes their work practices. For example, in the case of our vignettes, a planner without CareShifts performs all 6 sub-tasks. While she would use a computer to save data or calculate times, she would not have the feeling that the computer does the planning. In contrast, with CareShifts, in the supervisory paradigm, the human takes the output of the AI and overwrites it if needed. In other words, the human is excluded from the practices that the AI establishes. This change can decrease job satisfaction if it eliminates the humans from performing tasks that could potentially be the source of meaning in their job [91]. This also hints at why humans might not feel responsible for the output of their collaboration with automated systems when they enter their workplace. Especially, in higher levels of automation, the tasks performed by machines become more complex. Holding the human as the responsible supervisor of machines that are essentially beyond situational understanding and simply based on black-box models is impractical and may be unethical in itself. Our results reflected this as well. In the Supervisory (and Advisory) conditions, participants believed that both human and AI should be held accountable, explaining the outcomes and being praised or penalized for them, although they knew that the AI would not feel proud or embarrassed when facing the consequences of their work. Nevertheless, in the Interactive condition, as their contribution increased, they found it reasonable to be the explaining partner and receive praise or penalty. While it is not yet clear what it means to hold an AI accountable, these results imply that people do not want to render account for decisions and actions that they are not involved in. Recently, researchers proposed solutions for transparent, explainable, and understandable AI [50, 112] to overcome this challenge. Increasing transparency enhances trust in automation and improves teamwork [19]. Nevertheless, it has the disadvantage of causing information overload for humans in complex systems and impairs performance [93].

All in all, our results showed that in human-AI collaboration, people want to be directly involved in making decisions and performing tasks. Working with AI following the Interactive paradigm leads to higher job meaningfulness and satisfaction. In contrast, the lack of feeling of significance or competence induced through the Supervisory paradigm results in negative feelings and consequently job dissatisfaction. While some researchers believe that keeping the human in control through a supervisory role makes the design of the AI "human-centered" (e.g., [99]), our results showed that the more people feel that the AI is their teammate, the more content they are. These studies mainly associate the increase in human control with being human-centered. Previous literature on meaningful interaction with AI, however, has shown that other constructs, such as the feeling of competence through exercising skills, social relationships, and pursuing a purpose, can also play an essential role in human satisfaction and well-being in collaboration with AI [101]. In our study, the perception of autonomy and control was even higher when humans directly interacted with the AI rather than supervising it. A reason could be that delegating control does not necessarily imply autonomy, i.e., one can be in control of a task but does not have the autonomy in deciding how (or even whether) to conduct it. Furthermore, the positive correlation between the perception of AI as a teammate and job meaningfulness can confirm that people prefer to have a future of work with AI as a direct collaborator rather than a supervisor monitoring AI's actions.

These implications might raise the question: "If automation aims at increasing performance by reducing human errors, why should we strive to keep humans involved to maintain satisfaction?" The first answer is that due to the gradual advance of automated systems human involvement is required until full automation. Furthermore, previous research has shown a positive correlation between productivity (performance) and job satisfaction. Thus reduced job satisfaction can be considered an indicator or potential risk of decreased productivity [40, 75]. While aiming for performance goals in the short term might be beneficial, in the long term it can be detrimental. It is, therefore, vital to keep humans involved to reach both performance and well-being goals.

5.4 Implications for Design

5.4.1 Keep Humans Meaningfully Involved: Our results indicated that despite being the "final decision maker" in the Supervisory paradigm, people find their tasks less meaningful due to being excluded from their work practices. Unlike what the notion of these paradigm suggests, people feel a lack of autonomy and control as a result of delegating tasks to AI. Therefore, the first implication for the design of future human-AI collaboration is to keep humans involved in the actual work tasks. However, it is important to note that simply involving humans in collaboration is not enough. Our study revealed that although the Advisory paradigm involved humans in all sub-tasks, it did not result in a perceived sense of meaningfulness. This tackles the question of how this involvement should be designed. Previous work has shown that the distribution of tasks between human and AI collaborators has a significant effect on the perception of meaningfulness [91]. In situations where tasks are repetitive, uninteresting, and lack meaning, delegating such tasks to AI may actually be more beneficial and positively perceived by human workers. Hence, even when adopting the interactive paradigm, it is vital to assign tasks to humans that they see as a source of meaning. This can be achieved by addressing their psychological needs for competence, relatedness, and autonomy, as highlighted in previous research [30, 61, 62]. Nevertheless, it is important to acknowledge that performance is often the primary metric at work. Assigning specific tasks to humans might come at performance costs. For example, although it is crucial to prioritize the sense of autonomy and job meaningfulness for humans, there are situations where assigning tasks to AI might be more advantageous if it can substantially improve performance. It is essential to assess the unique characteristics of each task and find a balance between human involvement, job meaningfulness, and performance gains. This involves evaluating the trade-offs and determining the best allocation of tasks between humans and AI to achieve overall effectiveness and human worker satisfaction.

5.4.2 *Make AI a Teammate:* We found that the perception of AI as a teammate is positively correlated with job meaningfulness. But, how can AI be designed and introduced as a teammate? One of the main characteristics of a team is that the members work towards one or more common goals. It is, therefore, important to ensure that both human and AI have similar mental models of the goals, values, and roles in the team [55, 64]. One way to reach this is through providing explanations by AI about its actions, context, or role as a teammate [117]. Furthermore, improving the communication between human and AI simplifies the negotiation of goals and revealing intentions, and enhances mutual understandability and predictability [63]. In a team, members are brought together to perform relevant tasks with interdependencies regarding workflow, goals, and outcomes [64]. Our results showed that this can be done by designing more interactive paradigms where human and AI share tasks and are in dialogue. However, having distinct roles and responsibilities despite these interdependencies can enhance the perception of task identity and significance which are prominent factors in increasing job meaningfulness [47]. Trust is another crucial factor for the success of human-AI teams. AI reliability and consistency in performing its tasks can increase human co-workers' trust in its expertise. Furthermore, transparency and explainability in actions and decisions, and accountability for outcomes, reduce uncertainty and build trust between team members [120]. Finally, social interaction is known as the glue that bonds team members together. One way to do this is to learn from human-human interaction and design the interface and behavioral characters of these systems in accordance with well-known social interaction forms [32]. Previous CSCW works also show that people expect AI agents to act like humans when collaborating with them [55]. While considering AI as a tool could limit the team in complex collaborative tasks, such anthropomorphic designs might lead to expectations in human collaborators of these systems about their (social) communication capabilities that these systems cannot fulfill and can be more disruptive than helpful [105]. A recently proposed solution is to recognize AI's non-human characters and shape and define it as an "*Otherware*"–a counterpart with quasi-social relationships with us which may lead to new forms of relatedness, and the experience of mutual support [52]. However, social interaction at work is not only limited to working hours. The connections established among colleagues in the workplace can extend beyond work-related contexts, leading to enjoyable and meaningful experiences, such as friendships or interactions during company events. While existing research suggests that AI systems (robots) are more likely to become good colleagues than friends [80], additional investigation is needed to determine whether the design of AI collaborators can or should replicate these social experiences for individuals in the workplace.

5.4.3 Hold Humans Accountable Only When They Are Involved: Our study revealed that perception of accountability correlates with job meaningfulness in human-AI collaboration. However, designing meaningful accountability in such collaborations requires addressing responsibility and achievement gaps [27, 80]. In our study, we observed that individuals find it more reasonable to be held accountable for the outcomes of collaboration when they directly interact with AI systems rather than merely supervising them. This highlights the importance of active involvement in the decision-making and execution processes to create a sense of accountability. Furthermore, our results indicated that simply being involved in the collaboration by confirming AI decisions is insufficient for cultivating a feeling of accountability. It is essential to have a clear definition of responsibilities and the work processes. Excluding individuals from tasks assigned to AI or keeping them unaware of how those tasks are performed can create ambiguity, reduce their perception of accountability, and consequently contribute to a responsibility gap. To enhance ownership and accountability in human-AI collaboration, it is beneficial to aim for designing transparent and explainable AI collaborators [41]. Furthermore, when individuals are actively involved in identifiable tasks and have a clear view of their contributions, they become more certain about the recognition they receive in cases of positive outcomes. This recognition helps bridge the achievement gap and creates a sense of accountability. In this case feeling accountable does not only serve as an agreement for explaining outcomes to a forum, but also can serve as a means for fulfilling the psychological needs for popularity and competence at workspace [29].

5.5 Limitations

In this paper, we presented an experimental vignette study. Vignettes are used in both within-subject and between-subject design factorial survey approaches to investigate how situational characteristics influence attitudes or intentions (e.g., [7, 15, 86, 87]). In these studies, intentions serve as proxies for actual behavior in real-life scenarios [33]. However, this methodology has a number of limitations: (1) participants assess the practices in anticipated vignettes rather than real experiences (2) vignettes are limited in presenting all possible contextual factors within a practice, which might lead to minimizing the variability of the conditions, and (3) vignettes might miss conveying details that are relevant to the measure variables [57, 76]. An important question regarding vignettes is whether reported behavioral intentions accurately reflect and measure respondents' actual behavior in the described scenarios. This raises concerns about external validity (generalizability of results). Previous research has shown that high mundane realism (the degree to which the materials and procedures involved in an experiment are similar to events that occur in the real world) improves external validity [33]. In our experiment, we aimed to address this concern by presenting clear and detailed descriptions of scenarios in the vignettes. Our findings also revealed that participants found the scenarios highly relatable and easy to imagine. However, similar to many other controlled lab experiments (in comparison with field studies), there is a trade-off between ensuring internal and relative validity, and external validity. Further studies our required for exploring other factors that could impact the results. Nevertheless, despite the aforementioned limitations, this methodology enabled the possibility for experimental variation and control in a not-yet-existing practice.

The vignette scenario addressed the work meaningfulness of the individuals responsible for scheduling the shifts for nurses who are mostly in administrative roles and are not necessarily nurses or caregivers. Therefore, we did not set a pre-screening criterion for the occupation of the participants. As mentioned earlier, contextual factors such as the relationship between shift planners and nurses, or familiarity with the work culture and environment that were not covered in the vignettes can affect the perception of job meaningfulness and accountability. Nevertheless, we believe that our study still ensures internal validity (the observed effects are caused by the manipulation of independent variables). Further studies are needed to explore the influence of other variables to ensure further external validity.

In our study, the coordination of sub-tasks in the interactive paradigm is still quite static. In an ideal condition, the coordination of sub-tasks within this paradigm should be defined as more dynamic and without rigid boundaries [118]. Thus, considering our results, the dynamic context in real-world settings might further strengthen the interactive paradigm.

6 CONCLUSION

In this paper, we conducted an experimental vignette study to understand the effect of the collaboration paradigm in human-AI collaboration on the perception of job meaningfulness and accountability. We compared three collaboration paradigms: supervisory, advisory, and interactive across success and failure outcome conditions. Our results showed that while existing research on human-AI collaboration strongly insists on the supervisory control paradigm, the perception of job meaningfulness and satisfaction is significantly higher in the interactive paradigm. Furthermore, people prefer having AI as a teammate rather than supervising it. Finally, we found that to hold people accountable for the outcome of human-AI collaboration they should be involved directly in making the decisions and performing actions. In the future, further studies are required to explore the balance between job meaningfulness and performance and the design of AI as a teammate.

REFERENCES

[1] Hussein A Abbass. 2019. Social integration of artificial intelligence: functions, automation allocation logic and human-autonomy trust. *Cognitive Computation* 11, 2 (2019), 159–171.

[2] Herman Aguinis and Kyle J Bradley. 2014. Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational research methods* 17, 4 (2014), 351–371.

[3] Torbjörn Åkerstedt. 1990. Psychological and psychophysiological effects of shift work. *Scandinavian Journal of Work, Environment & Health* 1 (16 VL 1990), 67–73. https://doi.org/10.5271/sjweh.1819

[4] James E Allen, Curry I Guinn, and Eric Horvtz. 1999. Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications* 14, 5 (1999), 14–23.

[5] Kara A Arnold, Nick Turner, Julian Barling, E Kevin Kelloway, and Margaret C McKee. 2007. Transformational leadership and psychological well-being: the mediating role of meaningful work. *Journal of occupational health psychology* 12, 3 (2007), 193.

[6] Christiane Atzmüller and Peter M Steiner. 2010. Experimental vignette studies in survey research. *Methodology* (2010).

[7] Katrin Auspurg and Thomas Hinz. 2015. Multifactorial experiments in surveys. In *Experimente in den Sozialwissenschaften.* Nomos Verlagsgesellschaft mbH & Co. KG, 294–320.

[8] Catherine Bailey, Ruth Yeoman, Adrian Madden, Marc Thompson, and Gary Kerridge. 2019. A review of the empirical literature on meaningful work: Progress and research agenda. *Human Resource Development Review* 18, 1 (2019), 83–113.

[9] Sarah Bankins and Paul Formosa. 2023. The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics* (2023), 1–16.

[10] Nishani Bourmault and Michel Anteby. 2020. Unpacking the managerial blues: How expectations formed in the past carry into new jobs. *Organization Science* 31, 6 (2020), 1452–1474.

[11] Mark Bovens. 2007. Analysing and Assessing Accountability: A Conceptual Framework1. *European Law Journal* 13, 4 (2007), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0386.2007.00378.x

[12] Cynthia Breazeal, Jesse Gray, Guy Hoffman, and Matt Berlin. 2004. Social robots: Beyond tools to partners. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759).* IEEE, 551–556.

[13] Michael A. Campion, Troy V. Mumford, Frederick P. Morgeson, and Jennifer D. Nahrgang. 2005. Work redesign: Eight obstacles and opportunities. *Human Resource Management* 44, 4 (2005), 367–390. https://doi.org/10.1002/hrm.20080 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/hrm.20080

[14] Albert Camus. 1955. The absurdity of human existence. *The meaning of life* (1955), 94–100.

[15] Francis G Caro, Christine Yee, Samantha Levien, Alison S Gottlieb, Joachim Winter, Daniel L McFadden, and Teck H Ho. 2012. Choosing among residential options: Results of a vignette experiment. *Research on Aging* 34, 1 (2012), 3–33.

[16] Shari Caudron. 1997. The search for meaning at work. *Training & Development* 51, 9 (1997), 24–28.

[17] Neal Chalofsky. 2003. An emerging construct for meaningful work. *Human Resource Development International* 6, 1 (2003), 69–83.

[18] Gilad Chen, Bradley L Kirkman, Ruth Kanfer, Don Allen, and Benson Rosen. 2007. A multilevel study of leadership, empowerment, and performance in teams. *Journal of applied psychology* 92, 2 (2007), 331.

[19] Jessie Y.C. Chen, Michael J. Barnes, Anthony R. Selkowitz, and Kimberly Stowers. 2016. Effects of Agent Transparency on human-autonomy teaming effectiveness. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 001838–001843. https://doi.org/10.1109/SMC.2016.7844505

[20] Cary Coglianese and David Lehr. 2016. Regulating by robot: Administrative decision making in the machine-learning era. *Geo. LJ* 105 (2016), 1147.

[21] Philip R Cohen and Hector J Levesque. 1991. Confirmations and Joint Action.. In *IJCAI*. 951–959.

[22] Ravit Cohen-Meitar, Abraham Carmeli, and David A Waldman. 2009. Linking meaningfulness in the workplace to employee creativity: The intervening role of organizational identification and positive psychological experiences. *Creativity Research Journal* 21, 4 (2009), 361–375.

[23] MJ Colligan and RR Rosa. 1990. Shiftwork effects on social and family life. *Occupational medicine (Philadelphia, Pa.)* 5, 2 (1990), 315—322. http://europepmc.org/abstract/MED/2203160

[24] Stephen H Courtright, Gary R Thurgood, Greg L Stewart, and Abigail J Pierotti. 2015. Structural interdependence in teams: An integrative framework and meta-analysis. *Journal of Applied Psychology* 100, 6 (2015), 1825.

[25] Haydee M Cuevas, Stephen M Fiore, Barrett S Caldwell, and LAuRA StRAtER. 2007. Augmenting team cognition in human-automation teams performing in complex operational environments. *Aviation, space, and environmental medicine* 78, 5 (2007), B63–B70.

[26] ES Dahl. 2018. Appraising black-boxed technology: the positive prospects. *Philosophy & Technology* 31 (2018), 571–591.

[27] John Danaher and Sven Nyholm. 2021. Automation, work and the achievement gap. *AI and Ethics* 1, 3 (2021), 227–237.

[28] Nick Darlington. 2019. Shift Schedules: The Ultimate How-To Guide. https://www.7shifts.com/blog/shift-schedules-how-to-guide/. Accessed: 2022-05-15.

[29] Edward L Deci and Richard M Ryan. 1985. Conceptualizations of intrinsic motivation and self-determination. In *Intrinsic motivation and self-determination in human behavior*. Springer, 11–40.

[30] Edward L Deci and Richard M Ryan. 2012. Self-determination theory. (2012).

[31] Bryan J Dik and Ryan D Duffy. 2009. Calling and vocation at work: Definitions and prospects for research and practice. *The counseling psychologist* 37, 3 (2009), 424–450.

[32] Brian R Duffy. 2003. Anthropomorphism and the social robot. *Robotics and autonomous systems* 42, 3-4 (2003), 177–190.

[33] Stefanie Eifler and Knut Petzold. 2019. Validity aspects of vignette experiments: Expected "what-if" differences between reports of behavioral intentions and actual behavior. *Experimental methods in survey research: Techniques that combine random sampling with random assignment* (2019), 393–416.

[34] Mica R Endsley. 2017. From here to autonomy: lessons learned from human–automation research. *Human factors* 59, 1 (2017), 5–27.

[35] Connor Esterwood and Lionel P Robert. 2020. Human robot team design. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. 251–253.

[36] Nancy J Evans and Paul A Jarvis. 1980. Group cohesion: A review and reevaluation. *Small group behavior* 11, 4 (1980), 359–370.

[37] Mingming Fan, Xianyou Yang, TszTung Yu, Q Vera Liao, and Jian Zhao. 2022. Human-AI Collaboration for UX Evaluation: Effects of Explanation and Synchronization. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–32.

[38] Gerald R Ferris, James H Dulebohn, Dwight D Frink, Jane George-Falvy, Terence R Mitchell, and Linda M Matthews. 2009. Job and organizational characteristics, accountability, and employee influence. *Journal of Managerial Issues* (2009), 518–533.

[39] Fanny Ficuciello, Guglielmo Tamburrini, Alberto Arezzo, Luigi Villani, and Bruno Siciliano. 2019. Autonomy in surgical robots and its meaningful human control. *Paladyn, Journal of Behavioral Robotics* 10, 1 (2019), 30–43. https://doi.org/doi:10.1515/pjbr-2019-0002

[40] Nicole Forsgren, Margaret-Anne Storey, Chandra Maddila, Thomas Zimmermann, Brian Houck, and Jenna Butler. 2021. The SPACE of Developer Productivity: There's more to it than you think. *Queue* 19, 1 (2021), 20–48.

[41] Randy Goebel, Ajay Chander, Katharina Holzinger, Freddy Lecue, Zeynep Akata, Simone Stumpf, Peter Kieseberg, and Andreas Holzinger. 2018. Explainable AI: the new 42?. In *International cross-domain conference for machine*

learning and knowledge extraction. Springer, 295–303.
[42] Matthew C Gombolay, Reymundo A Gutierrez, Shanelle G Clarke, Giancarlo F Sturla, and Julie A Shah. 2015. Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots* 39, 3 (2015), 293–312.
[43] Adam M Grant and Sharon K Parker. 2009. 7 redesigning work design theories: the rise of relational and proactive perspectives. *Academy of Management annals* 3, 1 (2009), 317–375.
[44] Neal Gross and William E Martin. 1952. On group cohesiveness. *Amer. J. Sociology* 57, 6 (1952), 546–564.
[45] Johannes Gärtner, Philip Bohle, Anna Arlinghaus, Werner Schafhauser, and Magdalena Widl. 2018. Scheduling Matters – Some Potential Requirements for Future Rostering Competitions from a Practitioner's View. (2018), 10.
[46] J Richard Hackman. 1976. Group influences on individuals. *Handbook of industrial and organizational psychology* 1455, 1525 (1976), 1–37.
[47] J Richard Hackman and Greg R Oldham. 1974. The Job Diagnostic Survey: An instrument for the diagnosis of jobs and the evaluation of job redesign projects. (1974).
[48] J Richard Hackman and Greg R Oldham. 1975. Development of the job diagnostic survey. *Journal of Applied psychology* 60, 2 (1975), 159.
[49] J Richard Hackman and Greg R Oldham. 1976. Motivation through the design of work: Test of a theory. *Organizational behavior and human performance* 16, 2 (1976), 250–279.
[50] Hani Hagras. 2018. Toward human-understandable, explainable AI. *Computer* 51, 9 (2018), 28–36.
[51] Wayne C Harris, Peter A Hancock, Erik J Arthur, and JK Caird. 1995. Performance, workload, and fatigue changes associated with automation. *The International Journal of Aviation Psychology* 5, 2 (1995), 169–185.
[52] Marc Hassenzahl, Jan Borchers, Susanne Boll, Astrid Rosenthal-von der Pütten, and Volker Wulf. 2020. Otherware: how to best interact with autonomous systems. *Interactions* 28, 1 (2020), 54–57.
[53] Yugo Hayashi and Kosuke Wakabayashi. 2017. Can AI become reliable source to support human decision making in a court scene?. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 195–198.
[54] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.
[55] Ming-Tung Hong, Jesse Josua Benjamin, and Claudia Müller-Birn. 2018. Coordinating agents: Promoting shared situational awareness in collaborative sensemaking. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 217–220.
[56] Ziagul Hosseini, Sven Nyholm, Pascale M Le Blanc, Paul TY Preenen, and Evangelia Demerouti. 2023. Assessing the artificially intelligent workplace: an ethical framework for evaluating experimental technologies in workplace settings. *AI and Ethics* (2023), 1–13.
[57] Michael R Hyman and Susan D Steiner. 1996. The vignette method in business ethics research: Current uses, limitations, and recommendations. *Studies* 20, 100.0 (1996), 74–100.
[58] William A Kahn. 1990. Psychological conditions of personal engagement and disengagement at work. *Academy of management journal* 33, 4 (1990), 692–724.
[59] Deborah L Kellogg and Steven Walczak. 2007. Nurse scheduling: from academia to implementation or not? *Interfaces* 37, 4 (2007), 355–369.
[60] Bradley L Kirkman and Benson Rosen. 1999. Beyond self-management: Antecedents and consequences of team empowerment. *Academy of Management journal* 42, 1 (1999), 58–74.
[61] Holger Klapperich and Marc Hassenzahl. 2016. Hotzenplotz: reconciling automation with experience. In *Proceedings of the 9th Nordic conference on human-computer interaction.* 1–10.
[62] Holger Klapperich, Alarith Uhde, and Marc Hassenzahl. 2020. Designing everyday automation with well-being in mind. *Personal and Ubiquitous Computing* 24, 6 (2020), 1–17.
[63] Glen Klien, David D Woods, Jeffrey M Bradshaw, Robert R Hoffman, and Paul J Feltovich. 2004. Ten challenges for making automation a" team player" in joint human-agent activity. *IEEE Intelligent Systems* 19, 6 (2004), 91–95.
[64] Steve WJ Kozlowski and Daniel R Ilgen. 2006. Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest* 7, 3 (2006), 77–124.
[65] Janet Landman. 1987. Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin* 13, 4 (1987), 524–536.
[66] Matthias Laschke, Alarith Uhde, and Marc Hassenzahl. 2020. Positive Work Practices. Opportunities and Challenges in Designing Meaningful Work-related Technology. *arXiv preprint arXiv:2003.05533* (2020).
[67] Gabriel A León, Erin K Chiou, and Adam Wilkins. 2021. Accountability increases resource sharing: Effects of accountability on human and AI system performance. *International Journal of Human–Computer Interaction* 37, 5 (2021), 434–444.
[68] Joseph CR Licklider. 1960. Man-computer symbiosis. *IRE transactions on human factors in electronics* 1 (1960), 4–11.

[69] Marjolein Lips-Wiersma and Lani Morris. 2009. Discriminating between 'meaningful work'and the 'management of meaning'. *Journal of business ethics* 88, 3 (2009), 491–511.
[70] Alon Lisak, Raveh Harush, Tamar Icekson, and Sharon Harel. 2022. Team Interdependence as a Substitute for Empowering Leadership Contribution to Team Meaningfulness and Performance. *Frontiers in Psychology* 13 (2022), 637822–637822.
[71] Frank Martela. 2010. Meaningful work-An integrative model based on the human need for meaningfulness. (2010).
[72] Frank Martela. 2017. Meaningfulness as contribution. *The Southern Journal of Philosophy* 55, 2 (2017), 232–256.
[73] Frank Martela and Tapani JJ Riekki. 2018. Autonomy, competence, relatedness, and beneficence: A multicultural comparison of the four pathways to meaningful work. *Frontiers in psychology* 9 (2018), 1157.
[74] Kirsten Martin. 2022. Excerpt from Ethical Implications and Accountability of Algorithms. In *Ethics of Data and Analytics*. Auerbach Publications, 453–462.
[75] Christina Maslach and Michael P Leiter. 2008. Early predictors of job burnout and engagement. *Journal of applied psychology* 93, 3 (2008), 498.
[76] Louis S Matza, Katie D Stewart, Andrew J Lloyd, Donna Rowen, and John E Brazier. 2021. Vignette-Based Utilities: Usefulness, Limitations, and Methodological Recommendations. *Value in Health* 24, 6 (2021), 812–821.
[77] James McGrath and Ankur Gupta. 2018. Writing a moral code: algorithms for ethical reasoning by humans and machines. *Religions* 9, 8 (2018), 240.
[78] Richard Mulgan. 2000. 'Accountability': an ever-expanding concept? *Public administration* 78, 3 (2000), 555–573.
[79] Clifford Nass, BJ Fogg, and Youngme Moon. 1996. Can computers be teammates? *International Journal of Human-Computer Studies* 45, 6 (1996), 669–678.
[80] Sven Nyholm and Jilles Smids. 2020. Can a robot be a good colleague? *Science and engineering ethics* 26, 4 (2020), 2169–2188.
[81] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 30, 3 (2000), 286–297.
[82] Sanja Petrovic. 2019. "You have to get wet to learn how to swim" applied to bridging the gap between research into personnel scheduling and its implementation in practice. *Annals of Operations Research* 275, 1 (2019), 161–179.
[83] Michael G Pratt and Blake E Ashforth. 2003. Fostering meaningfulness in working and at work. Teoksessa KS Cameron, JE Dutton & RE Quinn (toim.) Positive organizational scholarship.
[84] Martin Ragot, Nicolas Martin, and Salomé Cojean. 2020. Ai-generated vs. human artworks. a perception bias towards artificial intelligence?. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–10.
[85] John Roberts and Robert Scapens. 1985. Accounting systems and systems of accountability—understanding accounting practices in their organisational contexts. *Accounting, organizations and society* 10, 4 (1985), 443–456.
[86] Peter H Rossi. 1979. 14. Vignette analysis: uncovering the normative structure of complex judgments. *Qualitative and quantitative social research: Papers in honor of Paul F. Lazarsfeld* 176 (1979).
[87] Peter H Rossi and Andy B Anderson. 1982. The factorial survey approach: An introduction. *Measuring social judgments: The factorial survey approach* (1982), 15–67.
[88] Brent D Rosso, Kathryn H Dekas, and Amy Wrzesniewski. 2010. On the meaning of work: A theoretical integration and review. *Research in organizational behavior* 30 (2010), 91–127.
[89] William B Rouse and William B Rouse. 1991. *Design for success: A human-centered approach to designing successful products and systems*. Vol. 2. Wiley-Interscience.
[90] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.
[91] Shadan Sadeghian and Marc Hassenzahl. 2022. The" Artificial" Colleague: Evaluation of Work Satisfaction in Collaboration with Non-human Coworkers. In *27th International Conference on Intelligent User Interfaces*. 27–35.
[92] Pouria Salehi, Erin K Chiou, and Adam Wilkins. 2018. Human-agent interactions: Does accountability matter in interactive control automation?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 62. SAGE Publications Sage CA: Los Angeles, CA, 1643–1647.
[93] Beau G Schelble, Christopher Flathmann, and Nathan McNeese. 2020. Towards meaningfully integrating human-autonomy teaming in applied settings. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. 149–156.
[94] Jean Scholtz. 2003. Theory and evaluation of human robot interactions. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*. IEEE, 10–pp.
[95] Axel Schulte, Diana Donath, and Douglas S Lange. 2016. Design patterns for human-cognitive agent teaming. In *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, 231–243.
[96] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*.

59–68.
[97] Smruti J Shah and James P Bliss. 2017. Does Accountability and an Automation Decision Aid's Reliability Affect Human Performance in a Visual Search Task?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, 183–187.
[98] Thomas B Sheridan. 2012. Human supervisory control. *Handbook of human factors and ergonomics* (2012), 990–1015.
[99] Ben Shneiderman. 2022. *Human-Centered AI.* Oxford University Press.
[100] Linda J Skitka, Kathleen Mosier, and Mark D Burdick. 2000. Accountability and automation bias. *International Journal of Human-Computer Studies* 52, 4 (2000), 701–717.
[101] Jilles Smids, Sven Nyholm, and Hannah Berkers. 2020. Robots in the Workplace: a Threat to—or Opportunity for—Meaningful Work? *Philosophy & Technology* 33, 3 (2020), 503–522.
[102] John R Sparks and Joseph A Schenk. 2001. Explaining the effects of transformational leadership: an investigation of the effects of higher-order motives in multilevel marketing organizations. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 22, 8 (2001), 849–869.
[103] Michael F Steger and Bryan J Dik. 2010. 11 Work as Meaning: Individual and Organizational Benefits of Engaging in Meaningful Work. (2010).
[104] Michael F Steger, Bryan J Dik, and Ryan D Duffy. 2012. Measuring meaningful work: The work and meaning inventory (WAMI). *Journal of career Assessment* 20, 3 (2012), 322–337.
[105] Hamish Tennent, Solace Shen, and Malte Jung. 2019. Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 133–142.
[106] Studs Terkel. 1974. *Working: People talk about what they do all day and how they feel about what they do.* The New Press.
[107] Peg Thoms, Jennifer J Dose, and Kimberly S Scott. 2002. Relationships between accountability, job satisfaction, and trust. *Human resource development quarterly* 13, 3 (2002), 307–323.
[108] Alarith Uhde, Matthias Laschke, and Marc Hassenzahl. 2021. Design and Appropriation of Computer-Supported Self-Scheduling Practices in Healthcare Shift Work. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 145 (apr 2021), 26 pages. https://doi.org/10.1145/3449219
[109] Alarith Uhde, Matthias Laschke, and Marc Hassenzahl. 2022. Experiential Benefits of Interactive Conflict Negotiation Practices in Computer-Supported Shift Planning. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction* (Canberra, ACT, Australia) *(OzCHI '22).* Association for Computing Machinery, New York, NY, USA, 180–192. https://doi.org/10.1145/3572921.3572927
[110] Alarith Uhde, Nadine Schlicker, Dieter P Wallach, and Marc Hassenzahl. 2020. Fairness and decision-making in collaborative shift scheduling systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13.
[111] George Ursachi, Ioana Alexandra Horodnic, and Adriana Zait. 2015. How reliable are measurement scales? External factors with indirect influence on reliability estimators. *Procedia Economics and Finance* 20 (2015), 679–686.
[112] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 2017. Transparent, explainable, and accountable AI for robotics. *Science robotics* 2, 6 (2017), eaan6080.
[113] James C Walliser, Ewart J de Visser, Eva Wiese, and Tyler H Shaw. 2019. Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making* 13, 4 (2019), 258–278.
[114] Joel S Warm, William N Dember, and Peter A Hancock. 2018. Vigilance and workload in automated systems. In *Automation and human performance: Theory and applications.* CRC Press, 183–200.
[115] Kelly D Wason, Michael J Polonsky, and Michael R Hyman. 2002. Designing vignette studies in marketing. *Australasian Marketing Journal (AMJ)* 10, 3 (2002), 41–58.
[116] Julika Welge and Marc Hassenzahl. 2016. Better than human: about the psychological superpowers of robots. In *International Conference on Social Robotics.* Springer, 993–1002.
[117] Ruchen Wen, Zhao Han, and Tom Williams. 2022. Teacher, Teammate, Subordinate, Friend: Generating Norm Violation Responses Grounded in Role-based Relational Norms.. In *HRI.* 353–362.
[118] Wout van Wezel, Julien Cegarra, and Jean-Michel Hoc. 2010. Allocating functions to human and algorithm in scheduling. In *Behavioral operations in planning and scheduling.* Springer, 339–370.
[119] Maranke Wieringa. 2020. What to Account for When Accounting for Algorithms: A Systematic Literature Review on Algorithmic Accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20).* Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3351095.3372833
[120] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*

(Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300509

[121] Sangseok You, Jeong-Hwan Kim, SangHyun Lee, Vineet Kamat, and Lionel P Robert Jr. 2018. Enhancing perceived safety in human–robot collaborative construction using immersive virtual environments. *Automation in Construction* 96 (2018), 161–170.

[122] Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, Victor R. Lesser, and Qiang Yang. 2018. Building Ethics into Artificial Intelligence. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 5527–5533. https://doi.org/10.24963/ijcai.2018/779

[123] Rui Zhang, Nathan J. McNeese, Guo Freeman, and Geoff Musick. 2021. "An Ideal Human": Expectations of AI Teammates in Human-AI Teaming. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 246 (jan 2021), 25 pages. https://doi.org/10.1145/3432945

Received January 2023; revised July 2023; accepted November 2023