Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-024-03962-x OPEN

From challenges to opportunities: navigating the human response to automated agents in the workplace

Ivan Đula [1,2,3]✉, Tabea Berberena[1,2,3]✉, Ksenia Keplinger [4,5]✉ & Maria Wirzberger [1,2,3,5]✉

Workers are increasingly embracing Artificial Intelligence (AI) to optimise various aspects of their operations in the workplace. While AI offers new opportunities, it also presents unintended challenges that they must carefully navigate. This paper aims to develop a deeper understanding of workers' experiences with interactions with automated agents (AA) in the workplace and provide actionable recommendations for organisational leaders to achieve positive outcomes. We propose and test a simulation model that quantifies and predicts workers' experiences with AA, shedding light on the interplay of diverse variables, such as workload, effort and trust. Our findings suggest that lower-efficiency AA might outperform higher-efficiency ones due to the constraining influence of trust on adoption rates. Additionally, we find that lower initial trust in AA could lead to increased usage in certain scenarios and that stronger emotional and social responses to the use of AA may foster greater trust but result in decreased AA utilisation. This interdisciplinary research blends a systems dynamics approach with management theories and psychological concepts, aiming to bridge existing gaps and foster the sustainable and effective implementation of AA in the workplace. Ultimately, our research endeavour contributes to advancing the field of human-AI interaction in the workplace.

[1] University of Stuttgart, Cluster of Excellence EXC 2075 "Data-Integrated Simulation Science", Stuttgart, Germany. [2] University of Stuttgart, Interchange Forum for Reflecting on Intelligent Systems (IRIS), Stuttgart, Germany. [3] University of Stuttgart, Department of Teaching and Learning with Intelligent Systems, Stuttgart, Germany. [4] Independent Research Group "Organizational Leadership and Diversity", Max Planck Institute for Intelligent Systems, Stuttgart, Germany. [5] These authors contributed equally: Ksenia Keplinger, Maria Wirzberger. ✉email: ivan.dula@ife.uni-stuttgart.de; tabea.berberena@iris.uni-stuttgart.de; kkeplinger@is.mpg.de; maria.wirzberger@iris.uni-stuttgart.de

Introduction

AI is rapidly transforming the nature of work by altering an individual's workload and common work tasks, as well as by stimulating new processes and practices. Given the rapidly changing work environment, leaders need to carefully balance new opportunities (e.g. improved productivity, time and cost savings) and unintended challenges brought by the use of AI, such as increased managerial control or exacerbated discrimination (Buolamwini, 2022; Crawford, 2021). Previous research suggests that workers' perceptions of AI range from rather positive to very negative expressions (Alberdi et al., 2009; Bankins et al., 2022; Lind, 2001), and after interacting with AI, some professionals even start doubting their expertise (Lebovitz et al., 2022). Thus, it can be a huge challenge for workers to successfully use newly implemented AI in their work environment (Benbya et al., 2020). For now, additional insights and resources to tackle this predicament effectively would be of value. Thus, we set out to investigate the following research question: 'How does the interplay of workload, effort and trust impact the workers' willingness to adopt AI in their work environment?'

The main purpose of this paper is to foster a better understanding of workers' interactions with automated agents (AA) in the work environment, as well as to develop actionable recommendations for managers on how to navigate potential benefits and downsides of AI's use to achieve positive outcomes. An AA can be defined as an encapsulated computational system enclosed within an environment, demonstrating adaptable and independent actions to fulfil its designated goals (Wooldridge and Jennings, 1995). While AA include a wide range of technologies and levels of complexity, ranging from computers, computer systems, machines, robots and algorithms, to AI systems (Chugunova and Sele, 2022), we focus on digital AA, such as algorithms, generative large language models like ChatGPT and DALL·E 3, and chatbots (Glikson and Woolley, 2020) engaged in sets of delegated tasks in the workplace (Falcone and Castelfranchi, 2001). Studying the interaction between workers and digital AA is crucial due to their growing integration across various work environments. Their versatility and widespread application in transportation, healthcare, customer service, sales and knowledge work underscore their potential impact on processes and decision-making in different workplaces (Jussupow et al., 2024; Vanneste and Puranam, 2024).

Taking advantage of higher processing speed (Haenssle et al., 2018), as well as improved justice perceptions in decision-making (Schlicker et al., 2021), AA are increasingly being deployed in work-related tasks. There are different types of tasks that an AA could be implemented for, ranging from generating ideas and plans, to solving problems and deciding issues, resolving conflicts of viewpoints or interests as well as resolving conflicts of power and executing performance tasks (McGrath, 1984). Given the diversity of task types, our paper focuses specifically on intellective tasks, as defined by Laughlin (1980). Intellective tasks are characterised by the requirement to find a demonstrably correct answer through invention, selection, or computation, which are likely to be particularly common in knowledge-based work contexts, such as research institutions, financial firms, consulting companies and technology development organisations, making our findings applicable across a broad variety of workplaces.

Our paper offers several important contributions. First, we respond to the urging for computational modelling to advance theoretical frameworks and better capture dynamic processes in management and organisational science (e.g. Kozlowski et al., 2013; Vancouver and Weinhardt, 2012). This involves using Vensim Professional 9.4.0 (Ventana Systems Inc., 2023) software to develop and test a system dynamics simulation model that quantifies, evaluates and predicts workers' experience with AA in a broader socio-economic context of human-AI interactions in the workplace. The mathematical model both captures the fundamental features of the system's structure and simulates workers' experiences under different initial conditions to examine possible behavioural patterns that can emerge from the interplay of balancing and reinforcing feedback loops in the model. In system dynamics, reinforcing and balancing feedback loops indicate the presence of reinforcing and balancing processes in the system, which either compound change in the case of a reinforcing feedback loop, or oppose change and seek equilibrium in the case of a balancing feedback loop. Reinforcing feedback loops tend to generate growth and collapse behaviours, while balancing feedback loops tend to generate growth-seeking behaviours (Kim, 1999). A simulation-based approach offers a structural viewpoint and an explanation of how variables like workload, effort and trust influence human behaviour and attitudes toward AA in the workplace. Our findings suggest that lower-efficiency AAs may outperform higher-efficiency AAs due to the constraining effect of trust on adoption, that low initial trust in AA may in some cases lead to higher adoption of AA, and that being more emotionally and socially responsive to AA can result in higher trust and lower adoption of AA.

Second, we use the simulation results to make recommendations on how managers can ensure a sustainable and effective implementation of AA for their employees. Using a mathematical simulation allows us to consider the dynamic relationships developing between humans and AA over time, helping managers anticipate both the positive and negative impacts of AA use. As individuals increase their use of AA in the future, it is important to clarify pathways leading to a positive impact on the individual level.

Finally, our paper tends to the necessity of an interdisciplinary approach by answering recent calls for interdisciplinary research on human-AI interaction in the workplace (Potočnik et al., 2023; Moore et al., 2012). Blending a systems dynamics approach with management theories and concepts from psychology helps to reconcile and integrate conflicting findings from previous research, as well as to advance knowledge in the domain of Collective Human-Machine Intelligence (COHUMAIN) introduced by Gupta et al. (2023).

Background

Conceptual model. To be able to leverage the benefits of newly introduced technologies in the workplace, individual employees need a sufficient level of technology acceptance. With the rise of technological advances in many professional sectors over past decades, research on factors fostering or hindering individual technology acceptance resulted in several established theoretical models, which subsequently served as a foundation for related investigations across different domains. Among these frameworks, we find the Theory of Reasoned Actions (TRA; Fishbein and Ajzen, 1975), the Technology Acceptance Model (TAM; Davis, 1989) and subsequent extensions (Venkatesh and Davis, 2000; Venkatesh and Bala, 2008), the Motivation Model (MM; Davis et al., 1992), the Theory of Planned Behaviour (TPB; Ajzen, 1991), the Combined TAM and TPB (C-TAM-TPB; Taylor and Todd, 1995), the Model of PC Utilization (MPCU; Thompson et al., 1991), the Innovation Diffusion Theory (IDT; Moore and Benbasat, 1991) and the Social Cognitive Theory (SCT; Compeau et al., 1999; Compeau and Higgins, 1995) to receive particular recognition. Integrating all of these frameworks based on comprehensive empirical research, Venkatesh et al. (2003) build the Unified Theory of Acceptance and Use of Technology (UTAUT), which specifies performance expectancy, effort expectancy, social

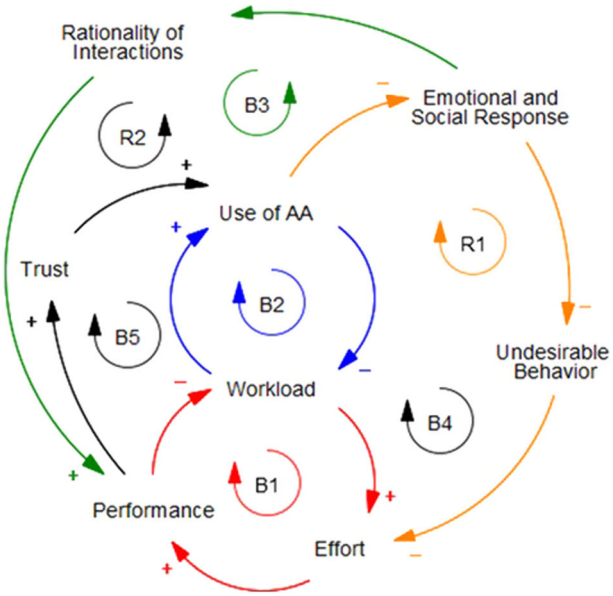

Fig. 1 Conceptual model of the use of AAs in organisations.

influence and facilitating conditions as core predictors for both the intention to use and the actual use of technical systems. These influences are moderated by individual and organisational factors, such as gender, age, experience with the respective technology as well as the voluntariness of use. More recent extensions of the UTAUT propose the inclusion of hedonic motivation, price value and habit as additional predictors (Venkatesh et al., 2012). Even though related research efforts could explain what brings people to use a particular technology, they do not investigate consequential behavioural dynamics of use in more detail. In particular, a systematic perspective on changes in users' cognitive and affective states, such as increased experience of workload or frustration, as well as potentially undesirable behavioural consequences such as stealing or absence from the workplace, is yet to be investigated. Additionally, while the scope of existing research spans a wide range of empirical work (Dwivedi et al., 2019), we observe a lack of simulation modelling results on proposed influencing factors and behavioural dynamics.

To address this research gap, Fig. 1 builds on recent work by Đula et al. (2023) and introduces a conceptual model that leverages a broader selection of variables to describe the potential arising dynamics arising from the use of AA in the workplace. Generally, this type of modelling approach captures how different system elements, or variables, are interrelated by using cause-and-effect linkages in the form of loops that can be reinforcing or balancing in nature (Kim, 1999). Based on previous research, the conceptual model depicts the link between feedback loops B1 and B2 (Balfe et al., 2015), which visualises the relationship between workload and the use of AA, as well as the relationship between workload, effort and performance. For example, the higher the workload, the more AA is being used to reduce workload and with it the need to increase the user's effort and thus compromise the overall performance (Corgnet et al., 2019). However, according to Chugunova and Sele (2022), the use of AA reduces the emotional and social response of the user, which can lead to undesirable behaviours, such as unethical or illegal behaviours (de Melo et al., 2016; Moore et al., 2012), which can be seen in the reinforcing feedback loop R1. On the other hand, in situations where increased emotions and social interactions are detrimental, the reduced emotional and social response has been shown to increase the rationality of human-AI interactions and improve performance (Chugunova and Sele, 2022), as shown in the balancing feedback loop B3. To successfully integrate AA into a workplace, research highlights the importance of trust (Glikson and Woolley, 2020; Parasuraman and Riley, 1997). Research suggests that the better and more reliable AA works, the more AA is used, and over time trust increases (Ullmann and Malle, 2017; Wang et al., 2016; Glikson and Woolley, 2020). This also means that users lose trust in a poorly performing AA and thus decrease their use of AA (De Visser et al., 2017, Hoff and Bashir, 2015). The variable of trust adds feedback loops B4 and B5, as well as R2 to the model.

Simulation model. To truly understand the behavioural dynamics resulting from the interplay of feedback loops identified in the conceptual model, we deploy the system dynamics modelling approach (Forrester, 1961) and develop a simulation model of the use of AA at work. System dynamics is a methodology and mathematical modelling technique used to frame, analyse, understand, discuss and solve complex issues and problems by focusing on the structure of physical, biological, or social systems as an endogenous driving force behind the behaviour of those systems (Lane, 1999). The central concepts of system dynamics are stocks, flows and feedback (Sterman, 2000, p. 191). Stocks are accumulations that characterise the state of a system and generate information upon which decisions and actions are based. They give systems inertia, provide them with memory and create delays by accumulating the difference between process inflows and outflows (Sterman, 2000). Flows are rates of changes in stocks. If stocks can be described as states of the system, then flows are transitions between different states. Feedbacks occur when system outputs are routed back as inputs as part of a chain of cause-and-effect that forms a loop (Ford, 2010). One of the most important capabilities of system dynamics, and the main reason why we utilise it to study the use of AA in organisations, is its ability to deal with natural and human systems with high levels of dynamic complexity. In contrast to combinatorial or detail complexity, dynamic complexity arises from the interaction among agents over time, and not just from the number of components in a system or the number of combinations one must consider in making a decision (Sterman, 2000).

Our guiding principle in constructing the model was to incorporate only the necessary components to capture the feedback loops identified in the conceptual model, while ensuring the replication of typically observed behaviours within the literature on human-AI interaction in work environments. We limited the applicability of our model to the specific case of intellective tasks in knowledge-based work contexts. In doing so, we aim to facilitate a better understanding of the model and clarify the relationship between the model structure and behavioural outcomes, avoiding over-generalisation. We sequentially constructed the model, building one feedback loop at a time with a reference to the conceptual model and existing literature. After incorporating each feedback loop, we ran intermediate simulations to ensure the model behaved reasonably and made adjustments to the structure when necessary. The complete structure of the model is shown in Fig. 2.

The conceptual model contains nine variables and seven feedback loops (five balancing and two reinforcing) (see Table 1). Starting with the B1 loop and considering that the meaning of the concepts, such as workload, effort and performance is highly dependent on the context (see e.g. Gopher and Donchin, 1986; Kahneman, 1973; Sonnentag and Frese, 2002), it was necessary to clearly define and quantify these concepts. Workload can be defined as an accumulation of tasks that the worker is required to process. Within the organisational behaviour literature, it can be compared to 'job demands', which are aspects of jobs requiring

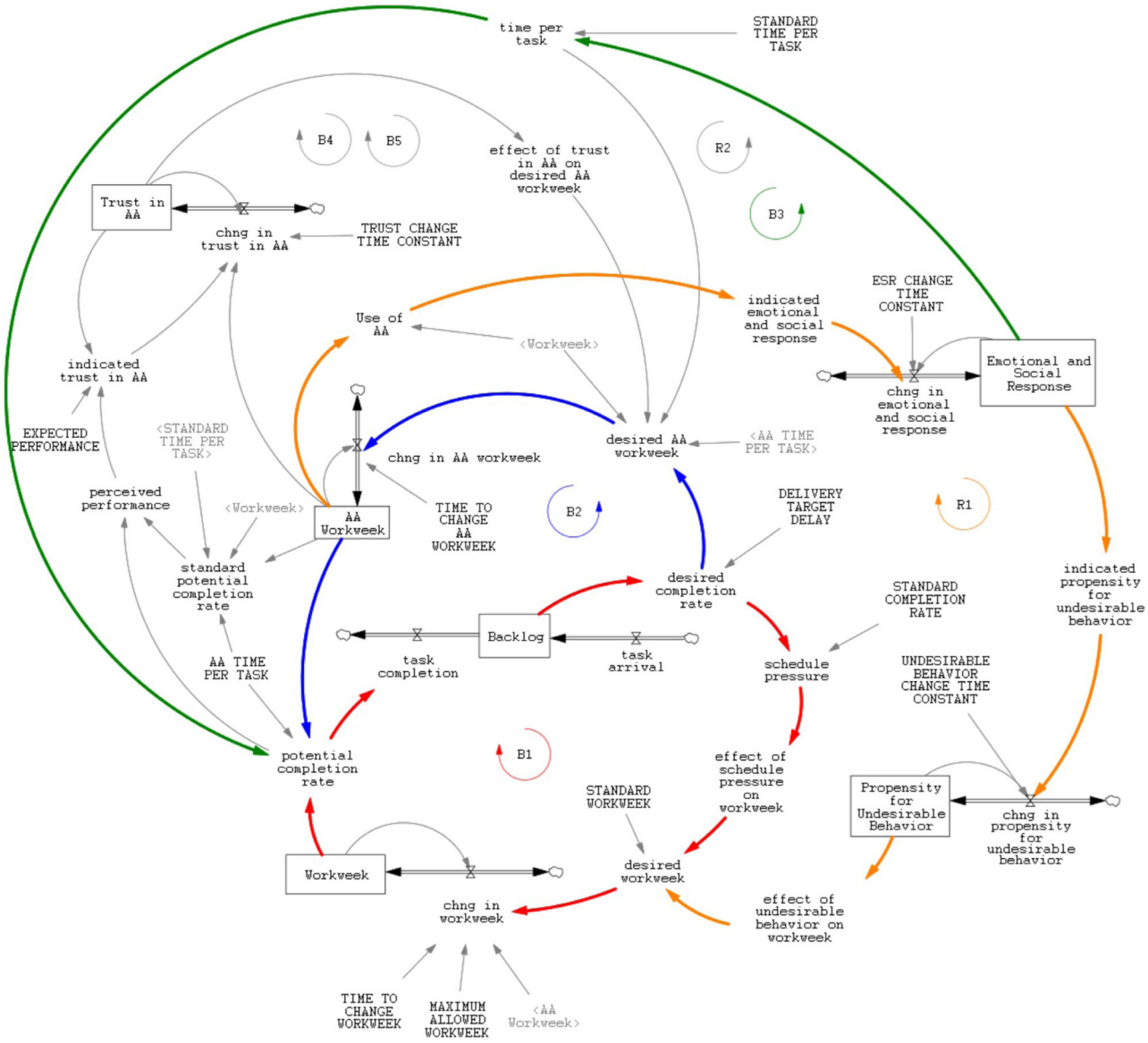

Fig. 2 Stock-and-flow structure of the simulation model.

Table 1 Causal links and polarities in feedback loops.

Loop	Causal relationships
B1	Workload → (+)Effort → (+)Performance → (-)Workload
B2	Workload → (+)Use of AA → (-)Workload
B3	Workload → (+)Use of AA → (-)Emotional and Social Response → (-)Rationality of Interactions → (+)Performance → (-)Workload
B4	Use of AA → (-)Emotional and Social Response → (-)Undesirable Behaviour → (-) Effort → (+)Performance → (+)Trust → (+)Use of AA
B5	Workload → (+)Effort → (+)Performance → (+)Trust → (+)Use of AA → (-)Workload
R1	Workload → (+)Use of AA → (-)Emotional and Social Response → (-)Undesirable Behaviour → (-)Effort → (+)Performance → (-)Workload
R2	Use of AA → (-)Emotional and Social Response → (-)Rationality of Interactions → (+)Performance→ (+)Trust → (+)Use of AA

sustained physical, emotional, or cognitive effort (Bakker et al., 2014) and corresponds to the NASA Task Load Index (NASA-TLX). Effort is defined as physical, emotional, or cognitive load that is actually allocated by the human agent to accommodate the demands imposed by the workload. Here we follow Paas et al. (2003) definition of cognitive load, which is seen as a product of task and subject characteristics. Finally, performance is defined as the rate at which the individual is able to process work-related tasks that directly serve their work goals. In that sense, it is synonymous with 'task performance' and 'in-role performance' concepts (see Motowildo and Van Scotter, 1994; Bakker et al., 2012).

In order to quantify these concepts, we decided to follow the well-established Homer (1985) model of worker burnout to structure relationships between workload, effort and performance. In Homer's model, the workload is depicted as backlog, modelled as a stock containing currently unresolved tasks; or, in our case example, lines of code that a software developer needs to

generate, translate, explain and verify. The backlog stock is changed through the inflow of new tasks and the outflow of completed tasks. Effort is quantified as a workweek, or the number of hours a worker spends independently working on solving tasks in a given week. The workweek is also a stock since it cannot be changed instantaneously. Instead, it changes over time, based on the worker's desired workweek. A desired workweek is a function of experienced schedule pressure, which is determined based on the desired and standard completion rate. If the desired completion rate falls below the standard, the worker experiences pressure to decrease their workweek, and if it is above the standard completion rate, the worker will experience pressure to increase their workweek. The workweek is directly related to the completion rate, which, in this model, corresponds to the performance variable.

We also decided to model the use of AA variable in the same way as the effort variable; with a stock indicating the number of hours per week a worker spends on solving tasks with the help of AA. Simply put, our model suggests that the worker can complete tasks on their own, as captured with the workweek stock, or by using AA, as captured with the AA workweek stock. The productivity of the human worker without and with the support of the AA is captured through time per task and AA time per task variables respectively. These two variables indicate how long on average it takes the human worker alone as well as the human worker using AA to complete a standard task. Together, they are used to determine the completion rate. The AA workweek also does not change instantaneously. There is a delay between the worker experiencing the desire to change the AA workweek and actually changing to the desired value. This delay is captured through the stock-and-flow structure incorporating the desired AA workweek as an input to calculate the required change in the AA workweek stock. The desired AA workweek depends on the desired completion rate and trust in AA. The desired completion rate is used to calculate how many hours of work are needed. If this value is above what the worker is allowed or willing to work, the value of the desired AA workweek will increase. In other words, if the number of tasks that need to be completed is higher than the worker can or wants to handle, the worker will use AA to handle that excess. This structure captures the B2 loop from the conceptual model.

Through the R1 loop, two more stocks are added to the model: the emotional and social response and the propensity for undesirable behaviour. In both cases, we are dealing with qualitative or so-called 'soft' variables that are difficult to quantify. Most modelling and simulation methods simply leave out such variables in order to increase the precision of their results. These kinds of variables, however, are important components of real-world systems and can strongly influence their behaviour. If we truly wish to understand dynamic social systems, it is necessary to find simple, explicit and sensible ways to model such variables. System dynamics is uniquely suited for this task as it offers a transparent way of approximating and quantifying these types of variables. In our model, we assume that emotional and social response, as well as the propensity for undesirable behaviour, can hold values between zero and one. For emotional and social response, zero indicates complete emotional and social disengagement of the worker from their organisational environment, while one indicates complete engagement. Similarly, a propensity of undesirable behaviour's value of zero indicates that the worker has no propensity to engage in undesirable behaviour at all, while one indicates the maximum possible propensity for undesirable behaviour. For the purpose of our paper and the specific context we are investigating, we have decided to keep the undesirable behaviour loop inactive. As Chugunova and Sele (2022) indicate, automation can be beneficial in contexts where emotions or social concerns are detrimental or harmful. We understand our specific context of a software developer focusing on intellective tasks to fall in this category, which is why we only activated the 'beneficial' feedback loop, namely the B3 loop. We decided to model this structure and provide an explanation for it in this section as it may be relevant in a different context as the aforementioned literature suggests.

The emotional and social response variable, following the conceptual model, depends on the use of AA variable, which we model as the share of AA workweek in total workweek (workweek and AA workweek combined). If this percentage increases, that means that the worker is spending more of their working hours using AA, which, with some delay, translates into decreased emotional and social response. The value of emotional and social response is used to calculate the propensity for undesirable behaviour, which once again adjusts with some delay. The main principle behind the equations is that if emotional and social response decreases, the propensity for undesirable behaviour starts increasing, which ultimately negatively affects the desired workweek. The relationships between the use of AA and emotional and social response, as well as between emotional and social response and propensity for undesirable behaviour, are captured through the so-called table functions, which we explain in the following section.

The B3 loop required us to establish the relationship between emotional and social response and the performance. We modelled this relationship by connecting the emotional and social response stock to time per task. If the stock decreases as a result of use of AA, the time required to complete a typical task reduces, making the worker more productive and increasing the completion rate. Mathematically, the relationship is modelled so that maximum emotional and social response results in time per task equal to standard time per task. A reduction in emotional and social response, i.e. from a maximum of 100 percent to 90 percent, is assumed to result in an equivalent reduction in time per task.

Finally, to capture the remaining components from the conceptual model, we added the trust in AA stock, which captures the level of trust a worker has in AA. Once again, this qualitative variable can assume values between zero and one, with zero indicating a complete distrust in AA, one indicating a complete trust, and the intermediate value of 0.5 indicating a neutral level of trust; neither trust or distrust. Trust in AA can change when AA is being used (AA workweek is above zero) and it is updated using the indicated value of trust, which is determined by the perceived performance and the expected performance. The perceived performance is a ratio of two different potential completion rates. One is the actual completion rate which uses current time per task to calculate the productivity of the worker, and the other is the standard potential completion rate which uses standard time per task. The expected performance captures what Glikson and Woolley, 2020 call 'features of virtual AI', such as visualisation and anthropomorphism, which may significantly impact the human user's expectations regarding AA's performance, while the actual performance of an AA moderates the direction of trust trajectory.

The rationale behind this structure is that using AA will increase workers' performance, and these performance gains will result in an increase in trust in AA. The difference between the performance when using AA and the performance if AA was not being used generates pressure toward the trust stock; however, this pressure is somewhat mitigated by worker's expectations. If the worker perceives an increase in performance when using AA, over time that will increase their trust in them (Diederich et al., 2022) as long as those performance gains are greater or equal to what the worker initially expected. Specifically, if the ratio is above the value of one, the worker has a positive perception of their performance and trust increases accordingly. If it is below one, the worker's perception is negative and trust decreases. On the other hand, if the perceived

performance is lower than the expected performance, the trust in AA will decrease even if the perceived performance when using AA is higher compared to performance without AA. The trust in AA stock subsequently influences the desired AA workweek variable. Increasing and decreasing the level of trust results in a proportional increase and decrease in the desired AA workweek. In the following section, we provide more detail regarding model equations and assumptions, as well as simulation settings.

Equations, core assumptions and initial simulation specifications. To simulate the model in Vensim professional software, it was necessary to insert mathematical equations for the variables, the values of initial stocks and parameters and non-linear relationships, and decide on simulation specifications. The equations we use are based on existing system dynamics models (e.g. Homer, 1985), or follow standard system dynamics modelling practices (e.g. Sterman, 2000). Values of stocks, parameters, non-linear relationships and the simulation specifications are selected with our scenario of a software developer using an algorithmic AA to complete intellective tasks, such as coding.

As stated before, the system dynamics approach to modelling and simulation is somewhat different compared to other methods, which typically prefer a much more empirical approach to model parameterisation. However, the main reason to use system dynamics is to generate insight into system behaviour, which is a result of its structure that contains difficult-to-quantify variables and to test policies that can potentially result in desired outcomes. We are particularly interested in learning about the nature of human-AI interaction in the workplace from a systems perspective, thus entering previously unexplored terrain and providing the necessary first step that future empirical work can build upon. In the rest of this section, we report on the most important equations and initial settings. The complete model listing following Rahmandad and Sterman (2012) guidelines for simulation-based research is available as supplementary material.

We devised a scenario in which every week five new coding tasks arrive in the developer's backlog and are processed with the expectation that it will take 1 week for them to be completed. The developer works a standard duration of 40 h per week, with the possibility to increase that amount to a maximum of 48 h per week if the desired workweek increases above the standard workweek. The desired workweek is determined by multiplying the standard workweek parameter with two effects—schedule pressure and undesirable behaviour. The backlog processes tasks as depicted in Eq. 1. The formulation for the workweek stock is shown in Eq. 2.

$$Backlog(t) = Backlog(0) + \int_0^t (task\ arrival(s) - task\ completion(s))\,ds;\ Backlog(0) = 5 \quad (1)$$

$$Workweek(t) = Workweek(0) + \int_0^t (chng\ in\ workweek(s))ds;\ Workweek(0) = 40 \quad (2)$$

Any work that the worker conducts using AA is captured in the AA workweek stock (see Eq. 3). Initially, this is set to zero and the scenario assumes that it will change only if the desired AA workweek increases, which will happen if the desired completion rate increases above what the worker is able to handle on their own and if the worker has a non-zero level of trust in AA.

$$AA\ Workweek(t) = AA\ Workweek(0) + \int_0^t (chng\ in\ AA\ workweek(s))ds;\ AA\ Workweek(0) = 0 \quad (3)$$

There are several important assumptions pertaining to the R1 and B3 loops in the model. First, the propensity for undesirable behaviour (Eq. 4) is initially set to zero and the emotional and social response (Eq. 5) is set to one. This means that our idealised scenario assumes that the worker is initially fully emotionally and socially engaged in the workplace and that they have no propensity for undesirable behaviour. Furthermore, the R1 loop contains two multiplier variables; one depicting the effect of the use of AA on emotional and social response and the other the effect of emotional and social response on the propensity for undesirable behaviour. The emotional and social response stock is used as a multiplier to calculate time per task. For example, we assume that reducing the emotional and social response to 0.9 (or 90 percent of the maximum value) reduces the time per task by an equivalent amount (10 percent). For the standard time per task, we assumed a value of 8 h per task, which, given the standard working week, indicates that the standard expectation in our imagined scenario is to complete one task every working day. Both stocks are adjusted with the assumed adjustment time of 4 weeks.

$$Propensity\ of\ Undesirable\ Behaviour(t) = Propensity\ for\ UndesirableBehaviour(0) + \int_0^t (chng\ in\ propensity\ for\ undesirable\ behaviour(s))ds;\ Propensity\ for\ Undesirable\ Behaviour(0) = 0 \quad (4)$$

$$Emotional\ and\ Social\ Response(t) = Emotional\ and\ Social\ Response(0) + \int_0^t (chng\ in\ emotional\ and\ social\ response(s))ds;\ Emotional\ and\ Social\ Response(0) = 1 \quad (5)$$

To finalise the model, the remaining equations and parameters related to the trust in AA are added, through which the remaining feedback loops (R2, B4 and B5) are translated into the simulation model. We initially assume that the worker is neutral towards the AA and set the initial value of the trust stock at 0.5 (see Eq. 6). The initial value of expected performance is set at 1, meaning that the worker has neutral expectations of AA's performance. The value of the trust stock is used as a multiplier to calculate the desired AA workweek. At full trust, the worker is assumed to prefer using the AA as much as possible to achieve the desired completion rate. This declines proportionally as the level of trust decreases from the maximum value. We assume that the adjustment time for trust spans 2 weeks. Based on Peng et al. (2023), we also assume that it takes 6.4 h to complete a task using the AA, compared to 8 h for the human worker to complete it on their own.

$$Trust\ in\ AA(t) = Trust\ in\ AA(0) + \int_0^t (chng\ in\ trust\ in\ AA(s)ds;\ Trust\ in\ AA(0) = 0.5 \quad (6)$$

Method

To evaluate the impact of the use of AA on workers' experiences, we conducted a series of simulation experiments using the model we developed in Vensim Professional 9.4.0 (Ventana Systems Inc., 2023). In all scenarios, we observe a simulation period of 20 weeks. We use a continuous simulation with a time step of 0.0625 weeks and the Euler integration method. Our focus is on backlog, workweek, AA workweek and trust in AA as indicators capturing the worker's experience. In addition to the base run (scenario 0) designed to show the system behaviour in equilibrium, we conduct four different simulation experiments to investigate the behaviour of the main variables.

In scenario 1, we create an external shock to the system by increasing task arrival from 5 to 7 in week 5. In scenario 2, in addition to the step increase in task arrival from scenario 1, we reduce the AA time per task from 6.4 to 3.2 h per task. We also conduct a sensitivity analysis to see how sensitive the focus indicators are to changes in AA time per task. In scenario 3, we again keep the step increase in task arrival and change the initial level of trust in AA to 0.75 and 0.25. We again conduct a

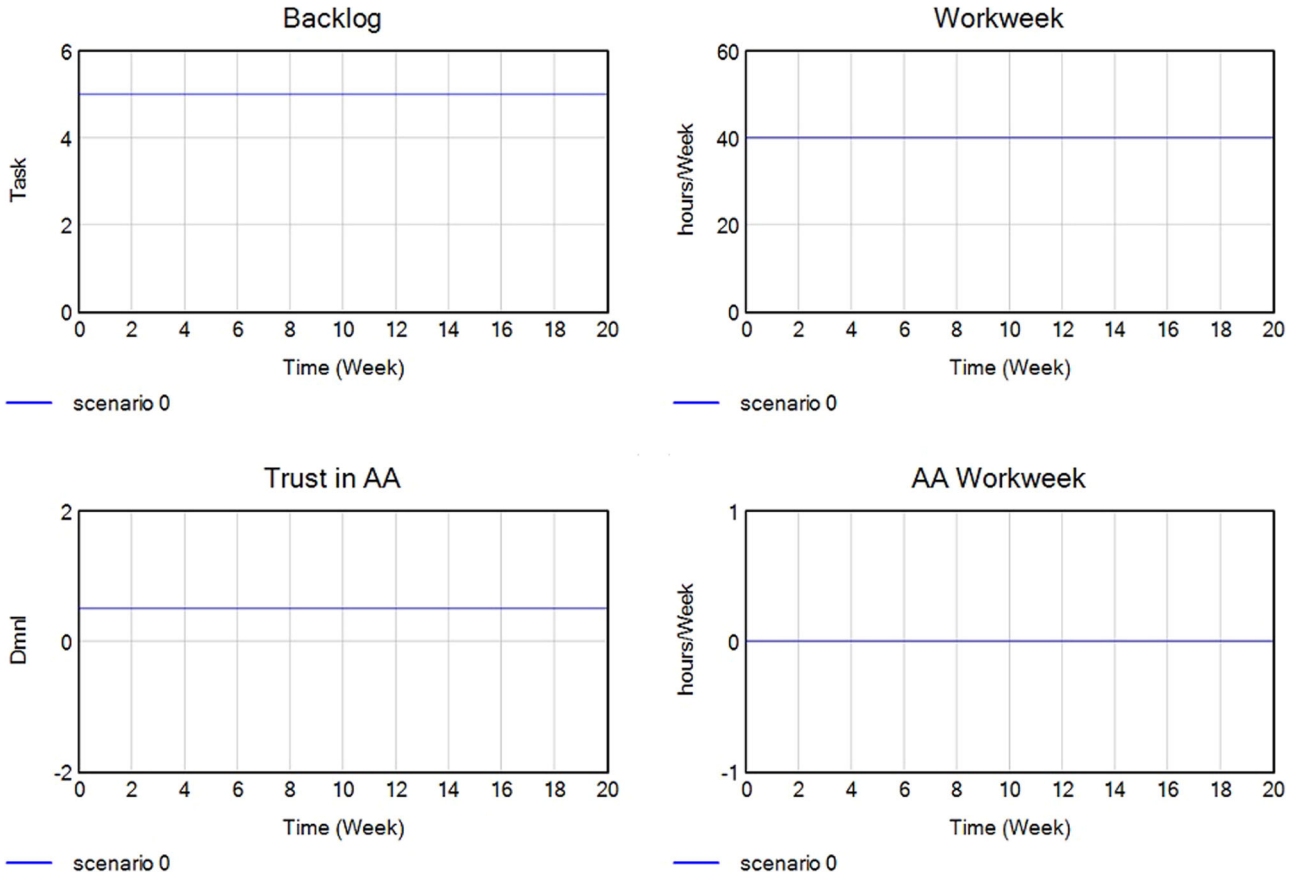

Fig. 3 Simulating a base run without external shocks. Without a change in workload, the worker does not change their use of AA and the system stays in equilibrium.

series of sensitivity analyses to examine how sensitive the AA workweek is to changes in trust adjustment time and how sensitive trust in AA is to expected performance under high and low initial trust scenarios. Finally, in scenario 4, whilst keeping the step increase in task arrival, we reduce the emotional and social response adjustment time from 4 to 1 week. In the following section, we provide details of each scenario run, graphs capturing the behaviour over time of selected variables and a short explanation of the structural causes of the observed behaviours.

Results

Scenario 0—Equilibrium. For the base run, we make no changes to the model equations. Simulation results for scenario 0 are presented in Fig. 3. Five new tasks are arriving into the backlog stock and five tasks are completed each week, which is why the backlog stock stays in stable equilibrium. There is no pressure on the worker to change the workweek or use AA. The worker's trust in AA has not changed since they do not use it. We use this simulation run as a technical check for the model. If there is no incentive to use the AA, the worker should refrain from using it and the model should stay in equilibrium, as demonstrated by the simulation run. For that reason, we will use results from scenario 1 as a reference point comparing all other scenarios.

Scenario 1—Step-up in workload. For scenario 1, we make one change to the base model. At week 5, there is a step increase in task arrival from 5 tasks per week to 7 tasks per week, and the workload stays at that value throughout the remaining simulation period. As shown in Fig. 4, there are several notable changes related to the worker's experience compared to the base run. The workload, as measured through the backlog stock, quickly increases due to the increased inflow of tasks ($M = 6.77$, $SD = 1.15$). This generates increased schedule pressure for the worker, which is alleviated by increasing the workweek ($M = 38.60$, $SD = 1.56$) and the AA workweek ($M = 6.42$, $SD = 4.33$). The workweek increases slightly following the step increase in tasks but quickly decreases below the initial value, as the worker begins to rely more on AA support. By week 20, the worker spends about 38 h per week without AA support and about 7 h with AA support. The backlog peaks around week 10 and settles at about 7 tasks per week by week 20. As long as the increased inflow of tasks continues, the trust in AA increases and the worker continues to use it for support. Over time, this brings further performance improvements through the B3 loop.

To summarise, increasing the workload results in a simultaneous increase in effort (workweek) and usage of AA (AA workweek). Due to the AA's effectiveness, the worker can quickly put the workload under control. However, because the number of tasks remains at an increased level throughout the simulated period, the worker is not able to completely stop using the AA. Continuous exposure to the error-free AA, however, increases their trust in AA and improves overall performance.

Scenario 2—A more efficient automated agent. For scenario 2, we keep the step increase in the workload (number of tasks) and adjust the AA time per task from 6.4 to 3.2. Essentially, we

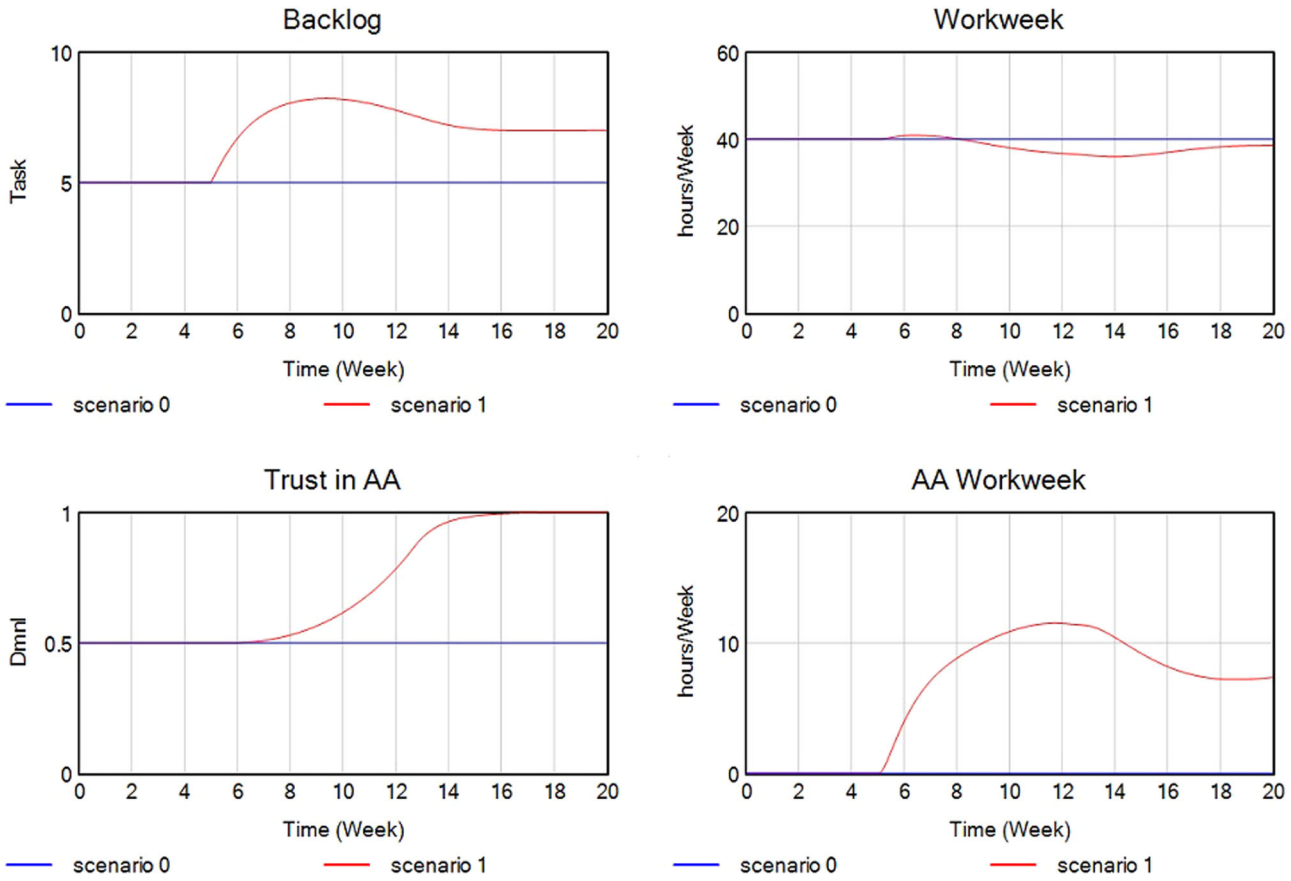

Fig. 4 Simulating a step increase in workload. In response to the increasing workload, the worker increases the use of AA. This, in turn, increases trust and eventually results in a decrease of total workweek compared to the baseline.

simulate the same external shock to the system as in scenario 1, but use an AA that is twice as efficient in completing tasks. Figure 5 shows the comparison of selected variables between scenarios 1 and 2. Several interesting and surprising results are worth noting. Intuitively, one might expect that a more efficient AA would result in a lower workload; however, the backlog stock exhibits similar behaviour in both scenarios and is, on average, even slightly higher in scenario 2 with the more efficient AA ($M = 6.87$, $SD = 1.18$). Trust in AA increases more slowly and ends up slightly lower in scenario 2 compared to scenario 1. The level of effort, as shown in the workweek stock, is also higher in scenario 2 ($M = 42.44$, $SD = 1.64$), while the use of AA, as shown in the AA workweek stock, is lower ($M = 2.72$, $SD = 1.74$). To sum up, increasing the number of tasks that the AA can complete, results in a higher workload, lower trust in AA, increased effort and reduces use of AA (see Fig. 5).

To test the reliability of this finding, we conducted additional sensitivity runs where we analysed the sensitivity of the main variables to the change in AA time per task. We tested for uniformly distributed values of this parameter between 0.1 h per task and 8 h per task, which equals the worker's time per task without the use of AA. The sensitivity runs show that increasing the AA efficiency (i.e. reducing the AA time per task) consistently leads to a higher backlog, lower trust in AA, increased workweek and reduced use of AA. In Fig. 6, we highlight the relationship between AA time per task and the backlog. Higher values of AA time per task, as shown in scenario 1 with a red line, lead to a lower backlog between weeks 12 and 20. Reducing the AA time per task results in an increasing backlog during that time period, with the lowest observed time of 0.1 yielding the highest backlog of about 8 tasks by the end.

Despite being counterintuitive, there are structural explanations behind these behaviours. Amore efficient AA reduces the need for frequent use to manage the backlog of tasks. As the backlog increases, the worker deploys the AA to deal with the increased workload. Once the AA resolves the backlog quickly, the worker reduces its use. This means that, compared to scenario 1, the worker interacts less intensively with the error-free AA, resulting in a slower increase in trust. The AA in scenario 1 is less effective in immediately resolving the initial problem of workload increase. However, this inefficiency results in more extensive interaction with the AA, exposing the worker to AI benefits more intensively. This exposure accelerates trust-building, increasing the worker's desire to use the AA even further. In scenario 2, due to the lower levels of trust, the worker relies more on their own effort and somewhat resists the use of AA.

Scenario 3—Initial trust. In scenario 3, we evaluate the impact of the initial value of trust and AA's performance expectations on the behaviour of the human-AI interaction system. In previous scenarios, we assumed that the worker starts with a neutral level of trust in AA and neutral expectations about AA's performance; however, this is likely not the case for many people. Certainly, there are individuals who are more receptive to the use of AA and those who are more resistant (Jussupow et al., 2024), along with different levels of expectations depending on AA's characteristics (Glikson and Woolley, 2020). For that reason, in addition to the step increase in tasks arrival, we ran two simulations with two different initial settings of trust; a value of 0.75 indicating high initial trust (scenario 3a) and a value of 0.25 indicating low initial

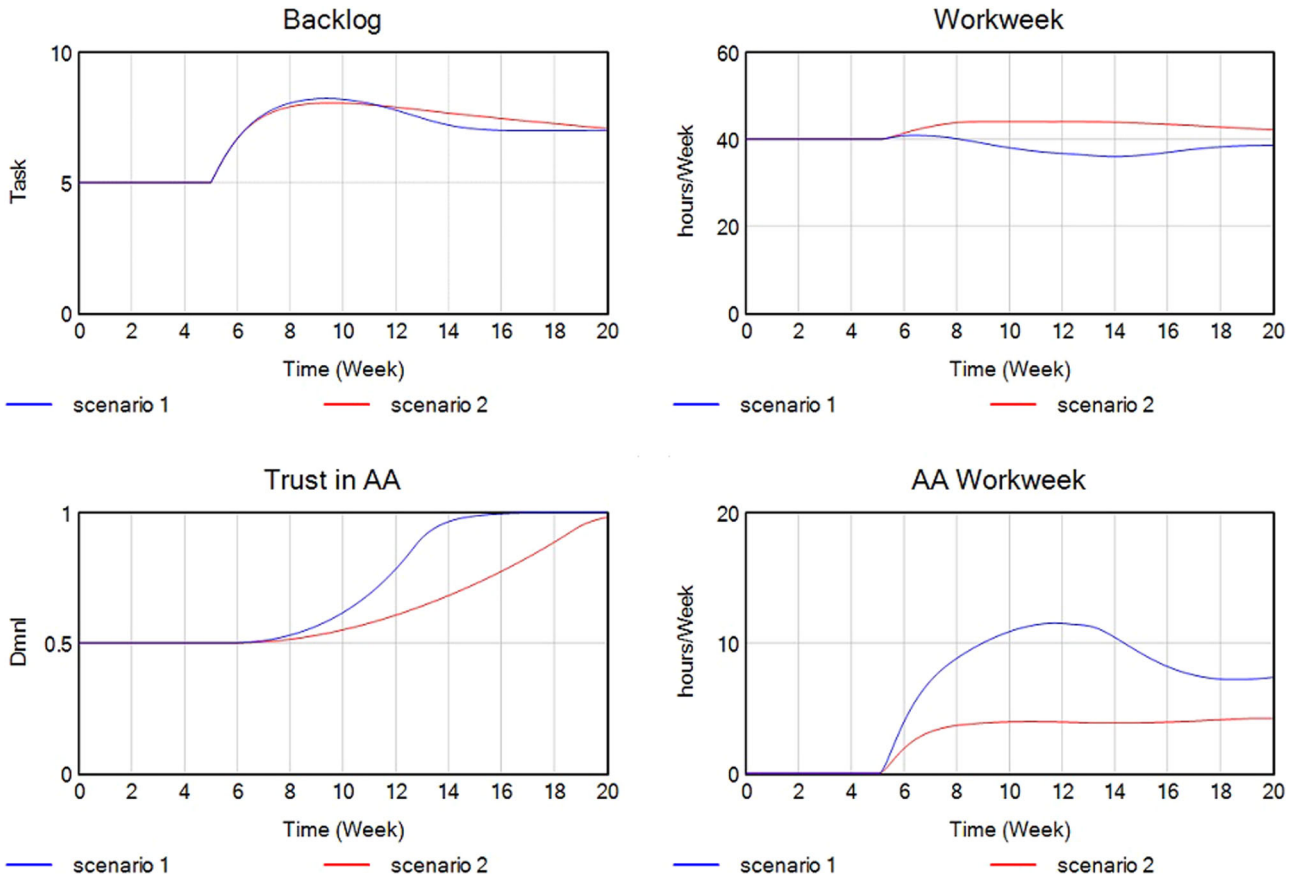

Fig. 5 Simulating an increase in AA productivity. A more productive AA results in fewer tasks being processed over time, as the worker relies on the AA less compared to the first scenario. Due to the higher effectiveness, the worker needs to use the AA less, which leads to less frequent interactions and a slower increase in trust.

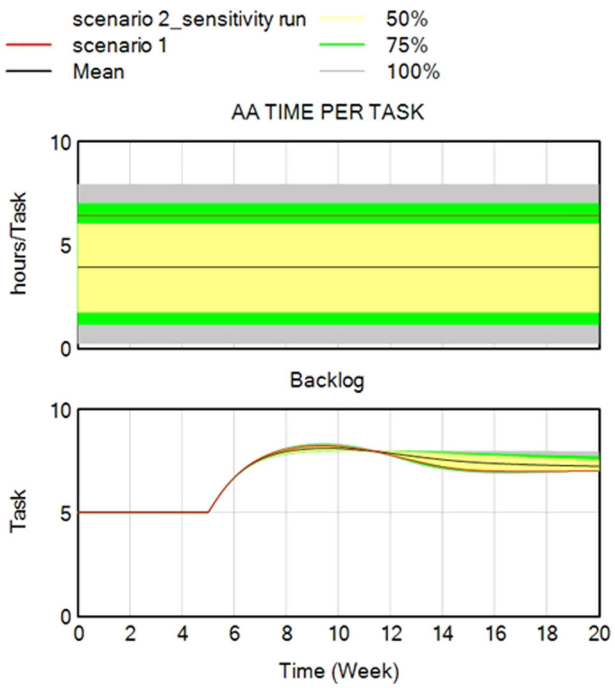

Fig. 6 Backlog sensitivity to change in AA time per task. Reducing the AA time per task consistently results in a higher backlog.

trust (scenario 3b). We present the simulation results compared to scenario 1 in Fig. 7.

As expected, in the high trust scenario, we observe the highest initial usage of AA and the lowest backlog ($M = 6.50$, $SD = 0.92$). Due to the initially high level of trust, the worker somewhat overuses the AA compared to scenario 1, especially following the step increase in workload ($M = 6.76$, $SD = 4.73$). By week 20, the workweek ($M = 37.75$, $SD = 2.31$) and other variables in the high trust scenario hold values similar to scenario 1, except for trust in AA, as it takes longer for trust in scenario 1 to reach that level. The low trust scenario, on the other hand, shows higher levels of both backlog ($M = 7.74$, $SD = 1.88$) and workweek ($M = 39.66$, $SD = 2.20$), as well as a slower, more gradual increase in the use of AA compared to the other two simulation runs ($M = 6.10$, $SD = 4.57$). These results corroborate previous findings regarding the difficulties of overcoming workers' aversion towards the use of AA (e.g. Jussupow et al., 2020; 2024). In scenarios 1, 3a and 3b, we assume an equally effective AA that can significantly reduce the worker's workload and neutral expectations of performance (the worker neither expects the increase or decrease in their performance). In scenarios 1 and 3a, this initially leads to an overshoot in AA use before settling to about 8 h per week. However, in scenario 3b, due to initial distrust towards the AA, there is no initial overshoot but instead, the use of AA slowly increases over time together with the trust, until the backlog is brought under control.

To gain more insight, we conducted sensitivity analyses for both the high-trust and the low-trust scenarios, where we tested

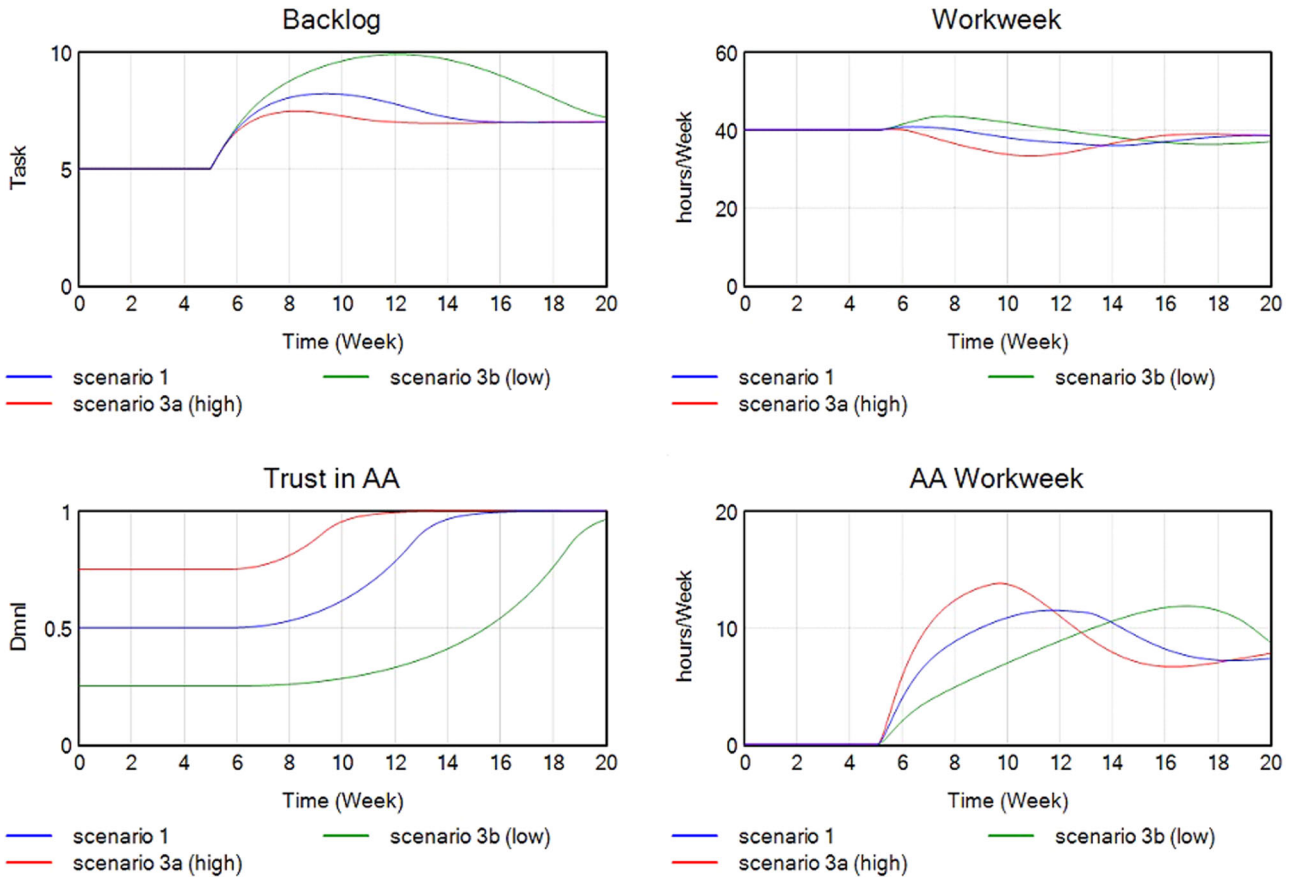

Fig. 7 Simulating changes in initial trust. Compared to the baseline, higher initial trust leads to a faster response to the increased workload, as indicated by a lower backlog and higher use of the AA. Over time, all scenarios converge to a similar level of AA usage.

how sensitive the AA workweek is to changes in trust adjustment time (scenarios 3a and 3b shown in Fig. 8), as well as how sensitive trust in AA is to expected performance (scenarios 3c and 3d shown in Fig. 9). We tested for uniformly distributed values of the trust adjustment time (0.1–10 weeks) and the expected performance (0.85–1.15) and observed that they can have a substantial effect on the main variables. Specifically, low initial trust in AA can lead to high initial adoption of AA if the worker's trust adjusts quickly enough. The effect appears to be much weaker when the initial trust in AA is already high, suggesting that the trajectory of AA use remains consistent in the high trust scenario regardless of how quickly the worker's trust adjusts.

In both the high and the low initial trust scenarios, we observe a strong impact of worker's expectations on the overall trajectory of trust in AA. Low expectations result in a faster increase in trust in AA, even in the low initial trust scenario. In contrast, high expectations can lead to a decrease of trust in AA, even when initial trust is high. This observation is crucial for the simulation model's overall validity as it demonstrates its ability to replicate empirically observed behaviours summarised by Glikson and Woolley, 2020 review of literature on human trust in AI. The review indicates that when initial expectations are high compared to the AI's perceived capability, trust in AI tends to decrease over time with increased interaction.

To summarise, the initial level of trust impacts the worker's experience as expected. Higher levels of trust lead to faster adoption and higher initial use of AA. Overall, there does not seem to be much difference between the high level of initial trust and the baseline, other than high trust scenarios seem to lead to much quicker reactions when opportunities to use AA arise. In contrast, low levels of trust, lead to resistance to the use of AA. Even when workers need to use AA to effectively manage their workload, it takes a long time to overcome the initial distrust. However, in the low-trust scenario, the worker avoids the initial overuse of AA and may end up using AA more compared to the higher-trust scenarios; all while maintaining higher levels of effort. The exception to this rule is a situation when the worker has an extremely short trust adjustment time. In those situations, the low trust scenarios result in even higher overshoot and usage of AA equivalent to high trust scenarios. Equally important are the performance expectations that the worker has about the AA. If these expectations exceed the perceived performance gains, trust in AA will decrease over time, resulting in lower-than-desired usage of AA. Conversely, if expectations are lower than the perceived performance gains, adoption speed, as well as the overall usage of AA, will be enhanced compared to a neutral scenario.

Scenario 4—Emotional and social responses. For the final scenario, we simulate the impact of increased emotional and social response sensitivity on the worker experience. To do that, we reduce the emotional and social response change time variable from 4 weeks to 1 week. This makes the worker far more susceptible to the use of AA, effectively strengthening the R1 and B3 loops. We keep the step increase in the workload (number of tasks) and use 0.5 as the initial value of trust in AA to make the simulation run comparable to scenario 1. The simulation results shown in Fig. 10 suggest that there are some similarities but also interesting differences between scenarios 1 and 4. The graphs of

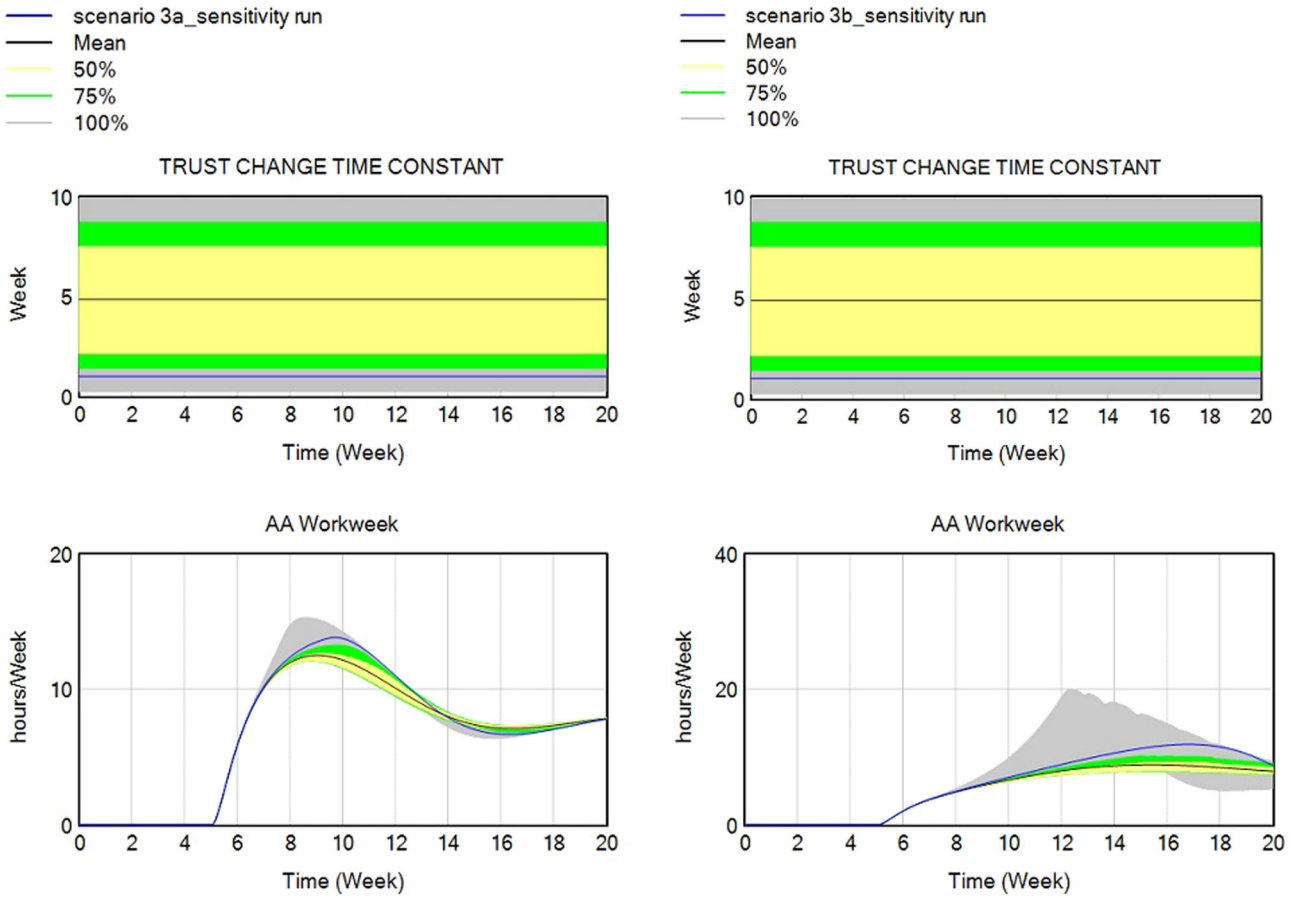

Fig. 8 Sensitivity of AA workweek to trust change constant in high- and low-trust scenarios. If the initial level of trust in AA is high, the trust adjustment time does not affect the use of AA to a great extent. If the initial level of trust is low, shorter trust adjustment times can result in a significant increase of AA usage.

backlog and workweek exhibit similar behaviours in both scenarios. The worker in scenario 4 has a slightly lower backlog ($M = 6.60$, $SD = 1.00$) and works more hours ($M = 39.30$, $SD = 1.16$) following the step increase, but by week 20 these values are identical to scenario 1. The reduction of change time for the worker's emotional and social response results, however, in a quicker increase of trust in AA compared to scenario 1. Despite that, the worker in scenario 4 uses AA significantly less than in scenario 1 following the step increase in tasks ($M = 5.49$; $SD = 3.52$).

These observations can be explained through the influence of the B3 feedback loop. By reducing the change time, the emotional and social response decreases faster in scenario 4. This results in a faster decrease in time per task, which further results in a faster increase in task completion rate. For that reason, the worker does not need to increase the AA workweek as much as in scenario 1. The gains in productivity due to the increase in rationality of interactions are sufficient to handle the increased workload. Ultimately, this may result in a worker using less and trusting the AA more compared to workers who are more resistant to the use of AA.

Discussion

The ability of AA to perform human-like cognitive tasks and their deployment in various workplaces are bound to have a profound impact on workers' experiences. To successfully deal with the transformational change of intellective work tasks, managers need to carefully balance the opportunities that come with the use of AA with risks that might emerge. Researchers in the field of human-AI interaction are slowly but surely filling the gaps in our knowledge related to these challenges, but many important questions remain unanswered. In this paper, we set out to deepen our understanding of human-AI interactions in work environments through the utilisation of system dynamics modelling and the use of a simulation-based approach to elicit simulation-based recommendations on how managers can ensure a sustainable and effective implementation of AA.

We first presented a literature-based conceptual model of the use of AA at work that conceptualised the dynamic complexity underneath human-AI interactions. Complex co-dependencies between different balancing and reinforcing feedback loops make it difficult to draw firm conclusions about possible behavioural outcomes of this system. For that reason, we developed a system dynamics simulation model and conducted a set of simulation experiments to study the effect of AA use on the workers' behaviour and experience. Several noteworthy observations emerged from our simulation work.

First, the primary motivation for the use of AA is externally caused by an increase in experienced workload. The worker uses an AA to keep up with the increasing demand for their labour. Simulations show that the AA, assuming significant performance advantages compared to the human worker, can indeed help alleviate the experienced workload pressure and stabilise the backlog. There are, however, potential unwanted and even problematic consequences. Scenario 1 shows that continuously relying on the use of AA builds trust. As the worker is

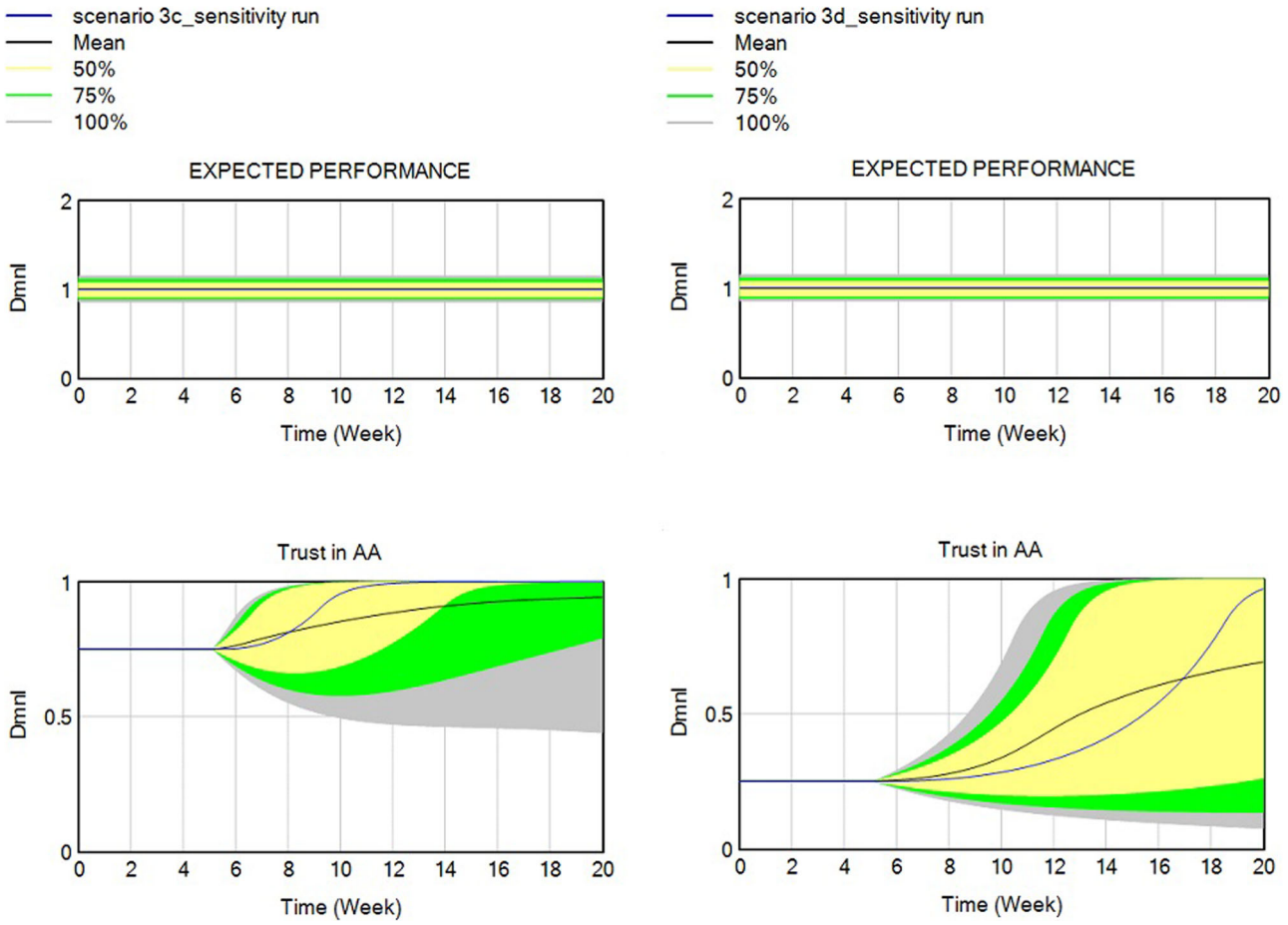

Fig. 9 Sensitivity of trust in AA to expected performance in high and low trust scenarios. Expectations have a strong impact on the development of trust in both scenarios. Low expectations consistently result in higher trust in AA. In contrast, high expectations lead to a slower increase in trust and, in some cases, even a decrease in trust in AA.

continuously exposed to beneficial outcomes resulting from the use of AA, they become more willing to take a back seat and allow the AA to permanently take over the excess of tasks. Interestingly, the AA does not simply take over and replace human labour. This process is gradual and requires a continuous growth of trust in AA, as well as an elevated level of workload. Managers should be aware of these dynamics and strive to mitigate them in the long run, either by reducing the workload back to the previous level, managing expectations about AA's performance, or by implementing policies aimed at detecting and suppressing potential undesirable behaviours.

One might think that a potential solution to any issue regarding the use of AA at work is to use a more effective AA capable of solving tasks much faster than humans or AAs from previous generations. Our simulations, however, suggest surprising and enlightening side-effects of that policy. There appears to be no significant workload reduction as a result of using a more efficient AA. If anything, it slightly underperforms compared to its less efficient counterpart. Moreover, the worker trusts the more efficient AA less and uses it less than the less efficient one, therefore ending up working more than in scenario 1. This is somewhat unexpected as intuitively one would expect that using a more efficient AA implies that workers will also use it more frequently. However, previous studies clearly show that trust is an important component of using AA in the workplace, and trust is built through exposure to AA. A less efficient AA means that the worker spends more time with it building trust. These results, as indicated in other studies and confirmed through our simulations, are contingent on expectations of performance held by the worker. If these expectations are higher than the perceived performance gains, the trust in AA will gradually erode over time with more interactions, even if the initial level of trust is high.

These results have implications for both managers and developers. For managers, the productivity of AA can be used to manage the effort level of their workers. On the one hand, highly productive AAs will result in reduced usage, yet, they will also foster higher effort and greater engagement by the workers. Less efficient AA, on the other hand, will be used more often, but they will also decrease the level of effort and engagement over time as workers begin to trust them more. For developers, this may have a profound impact on how they design AA, and we might already see some of these effects. The developers' objective is for their inventions to be used as much as possible. It stands to reason that they would not want to have the best and the most effective AA at work. Instead, AAs should be more effective at completing intellective tasks compared to their human counterparts, but not overly superior, so that workers still need to spend extensive time with them to get the desired outcome. Over the recent months, we observed many developers lock the premium features of their products behind paywalls, while simultaneously making the publicly available versions less capable compared to when they were released. Our model suggests that this may lead to an increased use of AA in the long run.

Another important aspect to consider is the initial level of trust in AA that the worker has. This can have a profound impact on the behaviour of the entire system. Unsurprisingly, higher levels

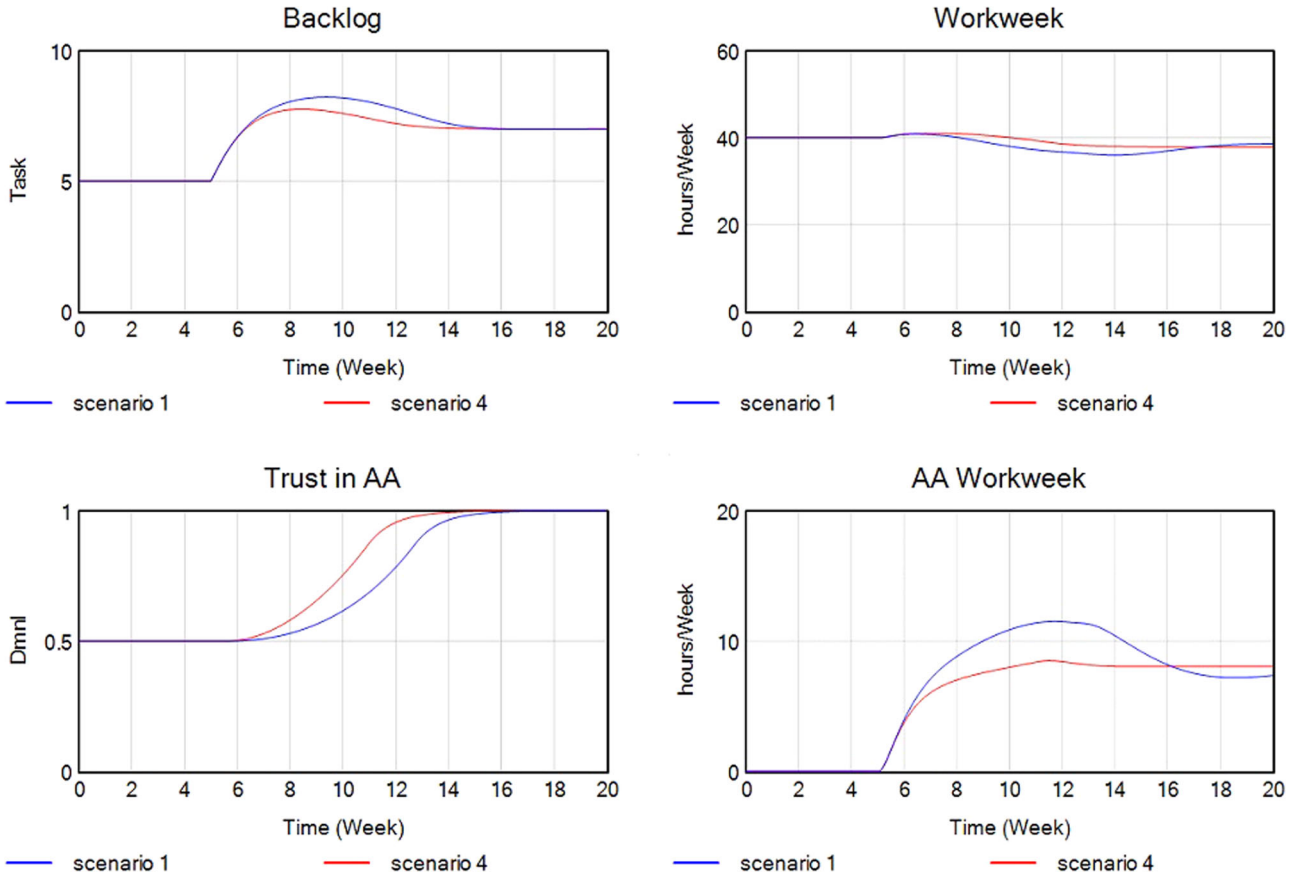

Fig. 10 Simulating a decrease in change time of emotional and social response. Increasing the worker's emotional and social sensitivity to the use of AA leads to a quicker increase in the worker's productivity, which, in turn, leads to better overall performance, while simultaneously reducing the worker's workweek and the use of AA compared to the baseline.

of trust lead to better performance and more usage of AA. Our model shows that the behaviour of workers dramatically changes when the initial trust is low and/or the performance expectations are high. This causes a vicious cycle of workers resisting the use of AA, therefore not being exposed to it, which leads to lower trust, which finally leads to further resistance to the use of AA. This dynamic can be somewhat overcome through the effectiveness of the AA. Regardless of the distrust in AA, at some point, it is not possible to ignore the benefits of their use. For managers, this indicates an absolute need to address their workers' trust in the AA they wish to introduce before the AA is actually implemented. Resistance to the use of AA can severely restrict their adoption and hinder their potential to improve the worker's performance. Our simulations show that, if the trust level can be changed quickly enough, it is possible to have a similar adoption to the high initial trust scenario.

Finally, our simulations showed that individual human characteristics may play an important role when it comes to how humans respond to AA in the workplace. Some people are adept at incorporating new technologies into their work. Instead of just using an AA, it becomes an essential part of their work life and they continuously find ways to get the most benefit out of it. In scenario 4, we simulate a situation, in which the worker's emotional and social response quickly changes based on the use of AA. We observe that in this scenario, the worker can reduce both their workload and the use of AA compared to scenario 1, while simultaneously having the same overall performance. For managers, this means that having workers who are extremely susceptible and adapt to the use of AA can result in some mild performance improvements, as they will be able to better utilise the AA; however, there is a risk that these workers will become increasingly better at finding ways to delegate their work to the AA and reduce their effort. Countering these opportunities should be a priority to avoid unwanted side effects.

While we could present an intriguing set of insights and related recommendations based on our simulations, we are aware that these need to be taken with a grain of salt. A conceptual model can only capture certain parts of the world and do so in a mostly idealised way. Hence, without validating our model predictions with empirical human data observed during human-AI interactions at work, we still lack an important piece of evidence. At the same time, by capturing dynamic mechanisms underneath interactions between humans and AA in a work environment in a highly formalised way, we lay the foundation for a thorough empirical investigation of related behavioural implications. Future steps in the scientific investigation should leverage these grounds to join forces with businesses and shed light on both workers' and AAs' behaviour under the above-outlined scenarios in naturalistic work environments.

Conclusion

Taken together, we apply evidence from system dynamics modelling to approach a highly relevant scenario in our tech-savvy society: Increasing exposure of humans in the workplace to the use of AA, such as chatbots or large language models, as supporting tools to complete work-related tasks. By introducing both reinforcing and balancing feedback loops of influencing factors, such as initial trust and emotional and social responses, we shed light on the effects of worker engagement, productivity and experienced

workload. Our findings suggest that lower-efficiency AA could outperform higher-efficiency ones, influenced by trust's impact on adoption rates. Additionally, low initial trust may accelerate AA adoption in specific scenarios, while individual characteristics significantly influence AA adoption effectiveness in work environments. Our approach allows us to put human-AI relationships under a computational microscope, providing fine-grained insights into the complexity of mechanisms underlying human-AI interactions in the workplace. As both researchers across different disciplines and practitioners (e.g. managers and software developers) will benefit from our work, we pave the way for productive human-AI partnerships in healthy work environments.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article [and/or] its supplementary materials. Access to the original model file and the full list of equations is provided through the following link: https://osf.io/5ktqs/?view_only=ca25d14597684297aaac79c87ce77516.

Received: 15 November 2023; Accepted: 14 October 2024;
Published online: 02 November 2024

References

Ajzen I (1991) The theory of planned behavior. Organ Behav Hum Decis Process 50(2):179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alberdi E, Strigini L, Povyakalo AA, Ayton P (2009) Why are people's decisions sometimes worse with computer support?. In: Buth B, Rabe G, Seyfarth T (eds) Computer Safety, Reliability, and Security. SAFECOMP 2009. Lecture Notes in Computer Science, vol 5775. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04468-7_3

Bakker A. B, Demerouti E, Sanz Vergel A. I (2014) Burnout and work engagement: the JD-R approach. Annu Rev Organ Psychol Organ Behav 3:389–411. https://doi.org/10.1146/annurev-orgpsych-031413-091235

Bakker A. B, Tims M, Derks D (2012) Proactive personality and job performance: the role of job craftingand work engagement. Hum Relat 65(10):1359–1378. https://doi.org/10.1177/0018726712453471

Balfe N, Sharples S, Wilson JR (2015) Impact of automation: measurement of performance, workload and behaviour in a complex control environment. Appl Erg 47:52–64. https://doi.org/10.1016/j.apergo.2014.08.002

Bankins S, Formosa P, Griep Y, Richards D (2022) AI decision making with dignity? Contrasting workers' justice perceptions of human and AI decision making in a human resource management context. Inf Syst Front 24:857–875. https://doi.org/10.1007/s10796-021-10223-8

Benbya H, Davenport TH, Pachidi S (2020) Artificial intelligence in organizations: current state and future opportunities MIS Q Exec 19(4):4. https://doi.org/10.2139/ssrn.3741983

Buolamwini J (2022) Facing the coded gaze with evocative audits and algorithmic audits. Unpublished PhD Dissertation, Massachusetts Institute of Technology, MA

Chugunova M, Sele D (2022) We and it: an interdisciplinary review of the experimental evidence on how humans interact with machines. J Behav Exp Econ 99:101897. https://doi.org/10.1016/j.socec.2022.101897

Compeau DR, Higgins CA (1995) Application of social cognitive theory to training for computer skills. Inf Syst Res 6(2):118–143. https://doi.org/10.1287/isre.6.2.118

Compeau DR, Higgins CA, Huff S (1999) Social cognitive theory and individual reactions to computing technology: a longitudinal study. MIS Q 23(2):145–158. https://doi.org/10.2307/249749

Corgnet B, Hernán-Gonzalez R, Mateo R (2019) Rac(g)e against the machine?: Social incentives when humans meet robots. GATE WP 1904—January 2019, Available at SSRN: https://ssrn.com/abstract=3324169

Crawford K (2021) The Atlas of AI: power, politics, and the planetary costs of artificial intelligence. Yale University Press

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 13(3):319–340. https://doi.org/10.2307/249008

Davis FD, Bagozzi RP, Warshaw PR (1992) Extrinsic and intrinsic motivation to use computers in the workplace. J Appl Soc Psychol 22(14):1111–1132. https://doi.org/10.1111/j.1559-1816.1992.tb00945.x

de Melo C, Marsella S, Gratch J (2016) People do not feel guilty about exploiting machines ACM Trans Comput Hum Interact 23(2):1–17. https://doi.org/10.1145/2890495

De Visser EJ, Montfort SS, Goodyear K et al. (2017) A little anthropomorphism goes a long way: effects of oxytocin on trust, compliance, and team performance with automated agents. Hum Factors 59(1):116–133. https://doi.org/10.1177/0018720816687205

Diederich S, Brendel AB, Morana S, Kolbe L (2022) On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research. J Assoc Inf Syst 23(1):96–138. https://doi.org/10.17705/1jais.00724

Đula I, Berberena T, Keplinger K, Wirzberger M (2023) Hooked on artificial agents: a systems thinking perspective. Front Behav Econ 2(2023):1223281. https://doi.org/10.3389/frbhe.2023.1223281

Dwivedi YK, Rana NP, Jeyaraj A et al. (2019) Reexamining the unified theory of acceptance and use of technology (UTAUT): towards a revised theoretical model. Inf Syst Front 21(3):719–734. https://doi.org/10.1007/s10796-017-9774-y

Falcone R, Castelfranchi C (2001) The human in the loop of a delegated agent: the theory of adjustable social autonomy. IEEE Trans Syst Man Cybern Part A Syst Hum 31(5):406–418. https://doi.org/10.1109/3468.952715

Fishbein M, Ajzen I (1975) Belief, attitude, intention and behavior: an introduction to theory and research. Addison-Wesley, Reading, Massachusetts

Ford A (2010) Modeling the environment, 2nd edn. Island Press

Forrester JW (1961) Industrial dynamics. The M.I.T. Press

Glikson E, Woolley AW (2020) Human trust in artificial intelligence: review of empirical research. Acad Manag Ann 14(2):627–660. https://doi.org/10.5465/annals.2018.0057

Gopher D, Donchin E (1986) Workload: An examination of the concept. In: Boff KR, Kaufman L, Thomas JP (eds) Handbook of perception and human performance, Cognitive processes and performance, vol. 2. John Wiley & Sons, p 1–49

Gupta P, Nguyen TN, Gonzalez C, Williams Wooley A (2023) Fostering collective intelligence in human-AI collaboration: laying the groundwork for COHUMAIN. Top Cogn Sci 1–28. https://doi.org/10.1111/tops.12679

Haenssle HA, Fink C, Schneiderbauer R et al. (2018) Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 29(8):1836–1842. https://doi.org/10.1093/annonc/mdy166

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

Homer JB (1985) Worker burnout: a dynamic model with implications for prevention and control. Syst Dyn Rev 1(1):42–62. https://doi.org/10.1002/sdr.4260010105

Jussupow E, Benbasat I, Heinzl A (2020) Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. In: Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference, June 15–17, 2020. https://aisel.aisnet.org/ecis2020_rp/168

Jussupow E, Benbasat I, Heinzl A (2024) An integrative perspective on algorithm aversion and appreciation in decision-making. MIS Quarterly. (In press). https://doi.org/10.25300/MISQ/2024/18512

Kahneman D (1973) Attention and effort. vol. 1063. Prentice-Hall, Englewood Cliffs, NJ

Kim DH (1999) Introduction to systems thinking. Pegasus Communications Inc, Waltham, Massachusetts

Kozlowski SWJ, Chao GT, Grand JA et al. (2013) Advancing multilevel research design: capturing the dynamics of emergence. Organ Res Methods 16(4):581–615. https://doi.org/10.1177/1094428113493119

Lane DC (1999) Social theory and system dynamics practice. Eur J Operational Res 113(3):501–527. https://doi.org/10.1016/S0377-2217(98)00192-1

Laughlin PR (1980) Social combination processes of cooperative problem-solving groups on verbal intellective tasks. In: Fishbein M (Ed.), Progress in social psychology, Hillsdale, New Jersey: Erlbaum, 127-155

Lebovitz S, Lifshitz-Assaf H, Levina N (2022) To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. Organ Sci 33(1):126–148. https://doi.org/10.1287/orsc.2021.1549

Lind E (2001) Fairness heuristic theory: Justice judgments as pivotal cognitions in organizational relations. In: Greenberg J, Cropanzano R (eds) Advances in organizational justice. Stanford University Press, Stanford, p 56–88

McGrath JE (1984) Groups: Interaction and performance. Prentice-Hall, Englewood Cliffs, New Jersey

Moore C, Detert JR, Treviño LK, Baker VL, Mayer DM (2012) Why employees do bad things: moral disengagement and unethical organizational behavior. Pers Psychol 65(1):1–48. https://doi.org/10.1111/j.1744-6570.2011.01237.x

Moore GC, Benbasat I (1991) Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res 2(3):192–222. https://doi.org/10.1287/isre.2.3.192

Motowidlo SJ, Van Scotter JR (1994) Evidence that task performance should be distinguished from contextual performance. J Appl Psychol 79(4):475–480. https://doi.org/10.1037/0021-9010.79.4.475
Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM (2003) Cognitive load measurement as a means to advance cognitive load theory. Educ Psychol 38(1):63–71. https://doi.org/10.1207/S15326985EP3801_8
Parasuraman R, Riley V (1997) Humans and automation: use, misuse, disuse, abuse. Hum Factors 39(2):230–253. https://doi.org/10.1518/001872097778543886
Peng S, Kalliamvakou E, Cihon P, Demirer M (2023) The impact of AI on developer productivity: evidence from github copilot. *arXiv preprint arXiv:2302.06590*. https://doi.org/10.48550/arXiv.2302.06590
Potočnik K, Chalmers D, Hunt R, Pachidi S, Townsend D (2023) Artificial intelligence: organizational possibilities and pitfalls. Journal of Management Studies, Call for Papers
Rahmandad H, Sterman JD (2012) Reporting guidelines for simulation-based research in social sciences. Syst Dyn Rev 28(4):396–411. https://doi.org/10.1002/sdr.1481
Schlicker N, Langer M, Ötting SK et al. (2021) What to expect from opening up 'black bloxes'? Comparing perceptions of justice between human and automated agents. Comput Hum Behav 122:106837. https://doi.org/10.1016/j.chb.2021.106837
Sonnentag S, Frese M (2002) Performance concepts and performance theory. Psychol Manag Individ Perform 23(1):3–25. https://doi.org/10.1002/0470013419.ch1
Sterman JD (2000) Business dynamics: systems thinking and modeling for a complex world, 51st print. McGraw-Hill, Irwin
Taylor S, Todd PA (1995) Understanding information technology usage: a test of competing models. Inf Syst Res 6(2):144–176. https://doi.org/10.1287/isre.6.2.144
Thompson RL, Higgins CA, Howell JM (1991) Personal computing: toward a conceptual model of utilization. MIS Q 15(1):125–143. https://doi.org/10.2307/249443
Ullmann D, Malle BF (2017) Human-robot trust: just a button press away. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI'17, (March 6–9), 309–310. https://doi.org/10.1145/3029798.3038423
Vancouver JB, Weinhardt JM (2012) Modeling the mind and the milieu: computational modeling for micro-level organizational researchers. Organ Res Methods 15(4):602–623. https://doi.org/10.1177/1094428112449655
Vanneste BS, Puranam P (2024) Artificial intelligence, trust, and perceptions of agency. Academy of Management Review. (In press). https://doi.org/10.5465/amr.2022.0041
Venkatesh V, Davis F (2000) A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 46(2):186–204. https://doi.org/10.1287/mnsc.46.2.186.11926
Venkatesh V, Bala H (2008) Technology acceptance model 3 and a research agenda on interventions. Decis Sci 39(2):273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x
Venkatesh V, Thong JYL, Xu X (2012) Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 36(1):157–178. https://doi.org/10.2307/41410412
Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27(3):425–478. https://doi.org/10.2307/30036540
Ventana Systems Inc. (2023). Vensim Professional 9.4.0 [Computer software] Available at: https://vensim.com
Wang W, Qiu L, Kim D, Benbasat I (2016) Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. Decis Support Syst, 86(2016):48–60. https://doi.org/10.1016/j.dss.2016.03.007
Wooldridge MJ, Jennings NR (1995) Intelligent agents: theory and practice. Knowl Eng Rev 10(2):115–152. https://doi.org/10.1017/S0269888900008122

Acknowledgements

Funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2075—390740016. We acknowledge the support by the Stuttgart Center for Simulation Science (SimTech).

Author contributions

ID: Conceptualisation, Resources, Methodology, Visualisation, Writing—original draft and Writing—review and editing. TB: Conceptualisation, Resources and Writing—review and editing. KK: Conceptualisation, Supervision, Project administration and Writing—review and editing. MW: Conceptualisation, Funding acquisition, Resources, Supervision and Writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical approval was not required as the study did not involve human participants.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Ivan Đula, Tabea Berberena, Ksenia Keplinger or Maria Wirzberger.