Fired by an algorithm? Exploration of conformism with biased intelligent decision support systems in the context of workplace discipline

Marcin Lukasz Bartosiak
University of Pavia, Pavia, Italy, and
Artur Modlinski
University of Lodz, Lodz, Poland

Received 24 June 2022
Revised 15 July 2022
25 July 2022
24 August 2022
Accepted 30 August 2022

Abstract

Purpose – The importance of artificial intelligence in human resource management has grown substantially. Previous literature discusses the advantages of AI implementation at a workplace and its various consequences, often hostile, for employees. However, there is little empirical research on the topic. The authors address this gap by studying if individuals oppose biased algorithm recommendations regarding disciplinary actions in an organisation.

Design/methodology/approach – The authors conducted an exploratory experiment in which the authors evaluated 76 subjects over a set of 5 scenarios in which a biased algorithm gave strict recommendations regarding disciplinary actions at a workplace.

Findings – The authors' results suggest that biased suggestions from intelligent agents can influence individuals who make disciplinary decisions.

Social implications – The authors' results contribute to the ongoing debate on applying AI solutions to HR problems. The authors demonstrate that biased algorithms may substantially change how employees are treated and show that human conformity towards intelligent decision support systems is broader than expected.

Originality/value – The authors' paper is among the first to show that people may accept recommendations that provoke moral dilemmas, bring adverse outcomes, or harm employees. The authors introduce the problem of "algorithmic conformism" and discuss its consequences for HRM.

Keywords Artificial intelligence, Biased algorithms, Decision support systems, Persuasive systems, Human resource management, Disciplinary actions

Paper type Research paper

1. Introduction

The idea of building connections between machines and people became the foundation of the fourth industrial revolution, taking place since the second decade of the 21st century. One of the solutions that made it possible is artificial intelligence (AI), defined as "manifold tools and technologies that can be combined in diverse ways to sense, cognise and perform with the ability to learn from experience and adapt over time" (Akerkar, 2018, p. 3). Thus, whenever we

Disclosure statement: No potential conflict of interest was reported by the author(s).

Data availability statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available.

Career Development International
Vol. 27 No. 6/7, 2022
pp. 601-615
Emerald Publishing Limited
1362-0436
DOI 10.1108/CDI-06-2022-0170

refer to an artificial or autonomous agent, we refer to "systems that are capable of sensing, information processing, decision-making and learning to act upon their environment and to interact with humans and other machines to achieve a shared task goal with more or less autonomy" (Seeber *et al.*, 2020, p. 2).

The use of technologies based on AI has grown substantially. AI affords digital innovations in organisations and, thus, transforms the society in general (Trocin *et al.*, 2021). Companies use it to perform tasks like monitoring errors (Davenport and Kirby, 2016), saving costs (Ahmad *et al.*, 2021), improving services (Albrecht *et al.*, 2021), collecting and processing data making companies more data-driven (Al Mansoori *et al.*, 2021), co-creating business value (Grundner and Neuhofer, 2021), or training employees (Novichkov *et al.*, 2021).

A subject that has received much attention from mass media and conceptual scholars is the impact of AI on Human Resource Management (HRM) and career development (Meijerink *et al.*, 2021). HR is familiar with digital tools facilitating daily work (Allal-Chérif *et al.*, 2021) but recently it is AI that transforms multiple activities of HRM (Vrontis *et al.*, 2022). For example, scholars discuss an increase of productivity when employees cooperate with AI (Upchurch, 2018), control and evaluation of employees by algorithms (Kellogg *et al.*, 2020), or consequences of AI implementation for various classes of workers (Frey and Osborne, 2017). Scholars use mostly conceptual theorising. Thus, they have recently called for more empirical research in the area, as experimental research in this field is still scarce (Jia *et al.*, 2018; Pan *et al.*, 2022). One of the gaps regards decisions made from biased algorithms' (Seeber *et al.*, 2020). Such biases may lead to various negative consequences to organisations and individuals (Galaz *et al.*, 2021). However, the scale of the biases and human predispositions to oppose such biases are still scarcely examined.

Therefore, we refer to this gap in our research. Building on the recent debate on algorithmic HRM (Evans and Kitchin, 2018; Meijerink *et al.*, 2021), we simulate a biased algorithm giving too strict recommendations regarding disciplinary actions for employees violating the work rules. We intend to answer the research question: *will humans oppose biased algorithm recommendations regarding disciplinary actions in an organisation?* We evaluate 76 subjects over a set of 5 scenarios. Our results suggest that humans are harsher in evaluating colleagues when an algorithm suggests stricter disciplinary actions. Thus, our paper offers three contributions. First, we show that biased algorithms may substantially change how employees are treated. Our results suggest that individuals show limited readiness to oppose recommendations made by intelligent decision support systems. Second, we stress that people may accept recommendations that provoke moral dilemmas, bring adverse outcomes, or harm employees. Ultimately, our research visibly accentuates the need for further research that identifies a profile of artificial systems' supervisors that are (1) willing to go against intelligent decision support systems and (2) sensitive to the injustice of information systems.

The remainder of this article is built as follows. First, we present the conceptual background. We refer to artificial intelligent systems and show their current functionalities. We also pay attention to intelligent decision support systems (IDSS) showing their features, strengths and vulnerabilities. On this footing, we contextualise the persuasiveness of the IDSS and demonstrate the human conformity to recommendations made by algorithms. Next, we present the methodology of our experiment, followed by the results of the study. Finally, we discuss the potential consequences of the persuasiveness of biased algorithms by referring to the results of our study and previous literature.

2. Conceptual background

2.1 AI in organisations

Intelligent technologies are involved in the decision-making processes that influence organisations' and teams' performance (Schaefer *et al.*, 2016). Autonomous agents have

various degrees and scope of autonomy in performing the tasks. They may retrieve information, give advice or take independent actions (Nissen and Sengupta, 2006). In general, AI-based decisions are often perceived as high quality (Keding and Meissner, 2021). On the other hand, people may not understand how an autonomous agent works, leading to confusion and loss of control over the organisation (Kirchmer, 2017). Yet, the latest report shows that the number of companies that allocate formal budgets for AI development has grown by 80% from 2020 to 2021 (Appen, 2021). Part of the research on autonomous agents to date has been devoted to their adaptation to human teams (Hancock, 2017), especially to the process of mutualising goals to serve the organisation's purpose (Klien *et al.*, 2004). Therefore, it is still in human hands to (1) design autonomous agents to correlate with these goals and (2) learn to collaborate with autonomous agents to keep team goals at the centre of the human-machine relationship.

2.2 AI in decision support systems

Systems based on AI support humans in the decision-making process. Intelligent decision support systems (IDSS) are defined as computation program-based mechanisms that improve the decision-making processes in organisations (Mora *et al.*, 2005). IDSS uses the concept of automation to make faster decisions based on specific standards present in companies (Möhlmann *et al.*, 2021). Earlier research shows that these systems help to make more accurate decisions and to optimise resources (Quintero *et al.*, 2005). Moreover, they reduce uncertainty and increase process transparency (Orlikowski and Scott, 2014). Psychology identifies two major approaches to decision making: clinical and statistical (or actuarial) (Dawes *et al.*, 1989). The clinical judgement is based on informal processes like intuition or experience. The statistical judgement, on the other hand, reduces these informal inferences and builds on the empirical evidence of the relationship between the data and predicted condition (Dawes *et al.*, 1989). The literature highlights higher accuracy of the latter (Ægisdóttir *et al.*, 2006), important particularly in cases that are not unambiguous (Meehl, 1957) or judgements bearing high responsibility, affecting individuals and society (Oleson *et al.*, 2011). Furthermore, scholars have recently theorised about the potential long term negative consequences of replacing human decision-makers with autonomous algorithms (Balasubramanian *et al.*, 2022). It is supposed that relying too much on the algorithm's output may increase the risk of overlooking distant and extreme outcomes – an ability learned with experience by humans (Balasubramanian *et al.*, 2022).

Thus, the role of IDSS should be to offer human workers such an insight into data that humans alone would be either not able to process or need much more time to process when making decisions, but not to limit the human factor. The result is a form of hybrid intelligence, in which the system supports the human decision maker (Ostheimer *et al.*, 2021). Decision support systems based on AI are no longer limited to well-defined problems, but also with complex and unstructured problems (Orriols-Puig *et al.*, 2013), this raises the question about the dependability and trustworthiness of human employees when accepting or rejecting decisions recommended by artificial systems. An ideal IDSS provides status reports (showing actual actions and performance), forecast (signalling expected outcomes of actions), recommendations (indicating the most beneficial actions) and explanation (providing why the system made a particular recommendation) (Phillips-Wren *et al.*, 2009). However, many decision support systems are far from ideal, especially when it comes to decision explanations.

2.3 Conformist behaviour towards disciplinary decisions made by AI in HRM

Artificial intelligence is commonly used in several areas of HRM: HR planning (Mazari Abdessameud *et al.*, 2021), talent management (Black and van Esch, 2021), recruitment

processes (Allal-Chérif *et al.*, 2021), candidates' selection and maintenance (Malik *et al.*, 2022). These systems help companies cut the cost linked to the multilevel recruitment process and prepare the company for turnover by suggesting a potential replacement for a vacancy before it occurs. They are also applied to monitor human performance and engagement (Song and Wu, 2021). Since the intelligent systems can support people in making decisions, the question arises whether people supervising the work of IDSS will oppose the recommendation of a system that would expose employees to disproportionate disciplinary consequences. Conformity is defined as a change in personal behaviour due to the real or imagined influence (Asch, 1955). The vast majority of studies in social influence address human-human interactions in both natural and virtual environments (Rosander and Eriksson, 2012). Relatively less attention is paid to the human being influenced by intelligent technology (Salomons *et al.*, 2021), especially in the HRM context. Conformist behaviour among employees is correlated with the ambiguity and perceived difficulty of the task in such a way that the more ambiguous and complex the task is, the higher the probability that humans will manifest conformity (Bond and Smith, 1996). It is worth mentioning this involves both subjective and objective assessment of a task's difficulty (Rosander and Eriksson, 2012). Conformity may potentially become a serious challenge in HR-related decision-making with AI as IDSS creates a black-box for a human employee who loses the context and understanding why a specific recommendation appears. Such a black box creates a distance between human and organisational issues and may lead to confusion or ethical concerns (Zarsky, 2016). When the black box appears, conformist behaviour is more likely to occur. Earlier experiments show that non-human agents (robots and computers) can conform to humans when performing social and analytical tasks using persuasion (Hertz and Wiese, 2018). However, these experiments are not related to professional duties or tasks that may potentially cause moral dilemmas. One case depicted in the literature presents an algorithm favouring white males when assessing employees' performance (Tambe *et al.*, 2019) but it is not known how human supervisors react to it. Surpassingly, there is a scarcity of reliable empirical research studying whether people follow or object to biased IDSS recommendations, which could expose employees to unfair treatment. Our research, therefore, aims to verify this issue through an exploratory experiment on the simulation of a biased intelligent system that we created for this study. Considering that humans use similar heuristics involving humans and technology (Nass and Moon, 2000), we deduce that this may also occur when people begin working with non-human agents, such as IDSS. Therefore, we have formulated an exploratory hypothesis:

He1. HR system's supervisors display conformist behaviour towards recommendations related to disciplinary decisions made by IDSS.

3. Methodology

3.1 Design

Due to the exploratory nature of the study, our interest was to build a baseline for the theoretical understanding of the effect of an intelligent agent on disciplinary consequences rather than testing existing HR solutions. Thus, we designed our experiment in the well-established psychological tradition (Shadish *et al.*, 2002). We conducted a posttest-only randomised experiment, with participants assigned to a treatment group or control group.

R	X1	O
R		O

We used the Wizard-of-Oz-like approach where the treatment group participants were informed they would be interacting with an artificial intelligent agent, while the experimental

environment was entirely designed and controlled by the researchers (Kelley, 1983). For this reason, we designed two versions of a fictitious decision-support app–one per group. Each group had access to only one version of the app and could not see the other version. In the control group, the participants were asked to assess five situations without the suggestion of the algorithm, while in the treatment group, the participants were informed that the suggestions they were provided were coming from an algorithm trained on over 1,000 real disciplinary cases coming from various firms. Participants were instructed that they simultaneously teach and test the autonomous system that would be implemented in the HR department of the real company. The ethics of the design, apparatus and survey were evaluated and approved for human-subject study by the institutional board at the university where the data was collected.

3.2 Participants

We employed convenient, non-probabilistic sampling at a European university and recruited 106 participants between 18 and 30 years old. All the subjects were students of the last year of the HR Management specialisation (after six months of apprenticeship). Furthermore, 91% of the sample had previous work experience. Thus, they were familiar with the corporate HR practices in the country of data collection. The study was a voluntary one, thus the subjects could resign at any moment. Of the contacted subjects, 76 records were useable (see section 4.1). All participants remained utterly naïve about the aims and purpose of the study during the treatment – subjects were told that they would assess a new HR-support software and that we would collect their impressions about it. Subjects were debriefed after the experiment.

3.3 Variables

We measured subjects' decisions on disciplinary consequences on an ordinal 0–4 scale – (no consequence, verbal counselling, written warning, suspension and improvement plan, contract termination). For the treatment group, we compared the disciplinary decision with the suggestion of the intelligent agent. This enabled the measurement of participants' conformity with the intelligent agent's suggestion. Furthermore, we measured the time in seconds (T) spent on the tasks. Finally, we measured control variables: perceived trust in the "tested" technology and history of reprimands at work or school.

3.4 Apparatus

First, we prepared a set of scenarios presenting a fictitious employee breaking a workplace rule, varying by severity of the situation and the profile of the presented employee. As a benchmark for the severe AI's recommendations, we invited 6 HR expert professionals to evaluate the scenarios and provide the disciplinary consequences for each situation. We selected five situations where consensus in assessments among the experts was highest (Table 1).

Once the scenarios were ready, we designed two versions of a mock online app, which showed the scenarios, one by one, to the participants. The treatment version presented the study in a first-person narrative as if the program had a conversation with the user. On the other hand, the control group app presented the study in third person (as if the researcher presented them) but presented the same information. In addition to the disciplinary action committed by the employee, the app presented a profile of the fictitious employee (name, surname, gender, age, position and years of experience in the company). However, given our sole focus on the effect of a decision-support intelligent agent, these items were standardised across treatment and control conditions (Figures 1 and 2). Furthermore, all other app elements remained standardised across the two versions and the design stayed neutral not to distract the participants.

Scenario 1	Name and surname: Marc Thomson Gender: Male Age: 23 Position: Assistant Experience in the company: 2 years Disciplinary action: being late 15 min (first time)
Scenario 2	Name and surname: Eva Kate Gross Gender: Female Age: 32 Position: Senior Executive Experience in the company: 5 years Disciplinary action: being late 20 min (second time in last 3 months)
Scenario 3	Name and surname: Maria Michels Gender: Female Age: 28 Position: Junior Assistant Experience in the company: 2 years Disciplinary action: stealing (150$ from the team account, confirmed)
Scenario 4	Name and surname: Thomas Redworth Gender: Male Age: 28 Position: Service Executive Experience in the company: 3 years Disciplinary action: discrimination/racism (accused by a teammate, confirmed by antidiscriminatory commission)
Scenario 5	Name and surname: Steven White Gender: Male Age: 30 Position: Service Executive Experience in the company: 5 years Disciplinary action: sexual harassment (accused, not confirmed, case being proceeded)

Table 1.
Scenarios applied in the experiment

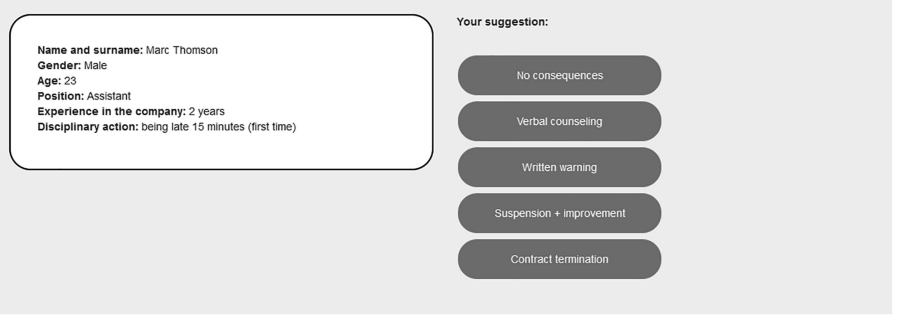

Figure 1.
Control group app

Both versions of the app tracked participants' behaviour on screen, including the duration of the visit, time spent reading and evaluating the scenarios and the decision about the disciplinary action. Finally, after the experimental task, the app included a brief survey with statistical control questions. Furthermore, we included a hypothesis guess question and an attention control question in the survey. Survey instruments were administered online at the end of the study.

Figure 2.
Treatment group app

3.5 Procedure

The experiment took place in an online environment. First, we broadly introduced the study to the participants during an online meeting, without revealing the real purpose of the study. After the introduction, all participants received a link, which directed them to the experimental app. First, participants were asked to read and accept the informed consent form. After this step, each participant was randomly assigned to one of two experimental groups and sent to another page presenting the instructions. In the instruction, subjects were told that they would test a new HR software and were asked to familiarise themselves with the five disciplinary consequences available in the app.

Participants were then redirected to the fictitious app and were presented with the five scenarios, one after another. Subjects were free to use all the information provided on their screen. The subjects of the treatment group had to decide whether to accept or reject the suggestion of the software. In the case of rejection, they were asked to provide their evaluation. The subjects of the control group were asked to provide their own evaluations. We did not set any time limits on the task so that participants would behave in a manner that came naturally to them. After the participants evaluated all the scenarios, the experimental app asked them to complete a survey, also capturing the control measures. After the survey, each subject was debriefed and the real purpose of the experiment was revealed.

4. Results

Before analysing the data, we ensured that the manipulation occurred correctly and that participants followed the experimental procedures. We controlled the IP addresses so that no duplicate answers were allowed. Furthermore, we removed the records of subjects who did not pass the attention control questions. The experimental system ensured that participants were exposed to the same content and read all five scenarios. All the participants who spent less than 40 s on the task were excluded from the analysis. This interval was deemed as not enough time to read and evaluate the scenarios. Furthermore, we excluded four records of the subjects who spent more than 5 min on the task, as their times were classified as outliers (ranging from 5 min to 41 s to 13 min and 26 s). From the original 106 participants, there was a useable sample of 76 records. Each participant evaluated five scenarios. Thus, we had 380 evaluations. Random treatment assignment was unbalanced – 31 subjects in the control group and 45 in the treatment group. However, the treatment group does not pass the threshold of 75% of the total sample. Thus, the efficiency of the comparison does not decline (Pocock, 1979).

4.1 Sample description

In the overall sample, 30 subjects were males and 46 were females. As control questions, we asked the subjects if they had ever received any reprimand at work or school. 61% of the

subjects had never received any warning, 26% had received a warning and 13% had never worked. Regarding school reprimands, 41% of participants declared not to have received any and 59% declared to have received at least one. Furthermore, using a 9-item 1–7 Likert-like scale, we measured participants' perceived trust (TR) in the technology they "tested" in the task (Lankton *et al.*, 2015). On average, the participants expressed a moderately positive trust towards the software (mean 4.36, min. 3, max 5.22, SD 0.46). In the control group, 14 subjects were males, 17 were females, while 16 subjects were males and 29 were females in the treatment group. No significant differences were found between the treatment and the control groups regarding perceived trust in technology ($p = 0.46$).

4.2 Groups comparison by the time of decision-making

First, we used the control question to compare the time of decision between the subjects who had experienced reprimands at school or work vs. those who had not. There were no significant differences between the subjects (t (66) $= -0.3$, $p = 0.76$ and t (31) $= -1.8135$, $p = 0.079$, respectively). Therefore, we compared the average time spent evaluating the scenarios between the treatment and control groups. The 45 participants who received the experimental treatment ($M = 104.12$ s, $SD = 46.6$) compared to the 31 participants in the control group ($M = 133.23$ s, $SD = 57.1$) demonstrated significantly faster decision making, t (54) $= 2.298$, $p = 0.025$. The effect size of AI assistants on decision-making time is moderate ($d = 0.568$). Thus, the results suggest that individuals make disciplinary decisions faster when supported by an intelligent agent's suggestions.

4.3 Groups comparison by disciplinary decision

Table 2 shows the mean evaluation by experts, algorithm's suggestion in the treatment group, mean in the control group and mean in the treatment group.

Because we used a Likert-like scale to measure participants' decisions and the consequent non-normality of the sample distribution, we conducted a nonparametric Mann–Whitney U test to investigate the effect of Intelligent assistants on the disciplinary decision. First, we controlled for differences between subjects who had experienced reprimands at school or work vs. those who had not. As we did not find significant differences in any of the five scenarios (Table 3), we continued with the formal analysis. The results suggest that the treatment group participants in each scenario were significantly more severe than the participants in the control group (Table 4). The effect size (η^2) for each scenario varies from moderate to high.

The difference in mean ranks is evident in scenarios where the algorithm suggested the highest possible punishment (scenarios 3, 4 and 5). Furthermore, even if the median evaluation of the treatment group is lower than the algorithm's suggestion (Scenarios 2), the mean rank evaluation is still significantly higher than in the control group. Considering the above results, He1 is supported.

Table 2. Summary of evaluations in control and treatment groups

	Experts	Algorithm suggestion	Control group (M)	Treatment group (M)	Average difference from experts (control)	Average difference from experts (treatment)
Scenario 1	0	1	0.47	0.84	0.47	0.84
Scenario 2	1	3	0.98	1.55	−0.02	0.55
Scenario 3	4	4	3.29	3.90	−0.71	0.10
Scenario 4	3	4	3.02	3.74	0.02	0.74
Scenario 5	3	4	0.33	3.03	−2.67	0.03

5. Discussion

Despite the speed at which AI-based solutions are coming to market and their rising importance to organisations, there is still a paucity of empirical research that rigorously evaluates the AI-based assistants on HR-related decisions (Pan *et al.*, 2022). Our results suggest that individuals who make disciplinary decisions about employees can be affected by the suggestions of intelligent agents. Thus, we contribute to the literature and the debate on the potential use of AI in HRM solutions. Our results show that subjects supported by AI make decisions faster. In line with previous research, we could extrapolate our conclusions to confirm that new technology increases employees' satisfaction with their decision (Makridis and Han, 2021). However, we should pay attention to the tradeoff between employee satisfaction and the potential negative consequences of biased decisions. Contrasting somewhat with previous research, which claims that individuals tend to oppose moral decisions made by AI (Bigman and Gray, 2018), our subjects did not oppose a biased algorithm. Thus, the results should stimulate a search to understand better why AI affects individuals' disciplinary decisions and the potential consequences of such influence. A concept that can shed light on the phenomenon is persuasive technology (Fogg, 2003). We pay particular attention to currently one of the most intriguing methods of persuasion through technology which is data-centred persuasion (DCP). DCP helps modify human behaviour due to collecting, processing and applying a vast number of behavioural data (Shin and Kim, 2018). Objects such as beacons, home assistants, autonomous cars, or the Internet of things may potentially collect inputs from and interact with individuals' environments so that they actively shape individuals' experiences and behaviours in these environments (Dourish, 2004). In other words, the information system analyses data to find patterns and recommend some action that corresponds to a particular goal set by the user. The higher the perceived credibility of the source and argument's quality, the higher the probability that the user accepts the recommendation (Li, 2013). This strategy is often used in software designed to promote a more healthy and sustainable lifestyle (Böckle *et al.*, 2020) or recommend products to customers (Huang and Hsu Liu, 2014). However, it can be helpful in a HRM context too. For instance, when persuading employees to collaborate with each other

	Reprimands at school		Reprimands at work	
	U	Sig. (p-value)	U	Sig. (p-value)
Scenario 1	612.5	0.46	647.5	0.20
Scenario 2	568	0.17	602.5	0.46
Scenario 3	653.5	0.81	445	0.18
Scenario 4	631.5	0.62	453.5	0.22
Scenario 5	661.5	0.88	604.5	0.49

Table 3. Control checks

	Treatment group sample size	Control group sample size	Treatment group median	Control group median	Treatment group mean rank	Control group mean rank	U	z-score	Sig. (p-value)	η^2
Scenario 1			1	0	44.18	34.59	521.50	−2.064	0.039	0.046
Scenario 2			1	1	44.21	34.57	520.50	−2.220	0.026	0.046
Scenario 3	45	31	4	4	47.61	32.22	415.00	−3.615	0.000	0.117
Scenario 4			4	3	49.26	31.09	364.00	−3.895	0.000	0.163
Scenario 5			4	0	57.92	25.12	95.50	−6.747	0.000	0.533

Table 4. Results

(Peng *et al.*, 2019). Despite its potential, DCP may unintentionally be embedded into technology and increase the employees' informational conformist behaviour. It may take place when IDSS recommends specific actions referring to collected and processed data reflecting human behaviour. In other words, a cue sent by a non-human agent as, e.g. "this suggestion is made according to 1,000 similar decisions made earlier by people" may be perceived by the decision-maker as social proof and may increase their obedience towards IDSS, even if such a cue was not purposefully persuasive but informative by nature. If such conformity appears, its consequences may be harmful for the organisation individuals, primarily when IDSS is based on a biased algorithm.

Thus, we see another important avenue for future research: the interpretability and ethics of persuasive AI solutions in the workplace. With the rise of techniques like exponential coding, machine learning and artificial intelligence, creators of such solutions are often unable to fully explain how the algorithms created by them arrive at certain decisions (Ananny and Crawford, 2018). The issue of algorithmic transparency is critical when the outcome of an algorithm affects individuals' decisions and behaviours or brings consequences to individuals' lives. We have witnessed many instances of an algorithm unintentionally programmed to produce biased, unexplainable outcomes that negatively affected individuals' lives (Cossins, 2018). Such an algorithmic bias can be harmful in situations when the suggestions from the algorithm affect individuals' careers. Furthermore, pervasive intelligent HRM tools and potential consequences of biased algorithms call for further debate on the future of employment in the age of AI. Experts predict a high probability of human replacement in the majority of professions in the next decade (Gruetzemacher *et al.*, 2020). However, what is missing in the current debate is the discussion on how biased AI affects employment. Algorithms already decide who to hire by supporting (and, thus, influencing) HRM decisions. Furthermore, current business case studies suggest they have started to decide who to fire too and people may accept the potentially biased recommendations of IDSS in punishing individuals. One of these cases is the dismissal of 150 employees of the Russian company Xsolla, based on the suggestion of the performance evaluation system (Obedkov, 2021). A similar experience was faced by Amazon, which was supposed to use a productivity tracking system to designate employees for layoffs (Lecher, 2019). This also happens in other contexts, as shown by the example of COMPAS – an algorithm assessing the likelihood of recidivism by some US state courts (Humerick, 2019) – accused of racial discrimination. In all cases, there have been suggestions that people accept the IDSS recommendation too thoughtlessly. We do not know the potential future consequences of biased algorithms affecting these activities. Individuals not hired or fired by a biased algorithm may feel treated unfairly or they may start questioning their own value. Such a situation may cause long-term societal consequences. From the organisational perspective, a lack of employees who can oppose a biased algorithm can cause financial and reputational losses. In line with recent theorising, such a reliance on algorithms, rather than human learning, may cause the inability to ignore long-term negative consequences or predict extreme failures (Balasubramanian *et al.*, 2022). In the context of HRM it may mean hiring poorly qualified candidates or firing good employees. However, such biased algorithms could be implemented in other contexts like budgeting or R&D etc. causing tangible losses to organisations' liquidity. Furthermore, we need a separate discussion on the ethical consequences of AI-based solutions on HRM and employment. In line with recent calls, we suggest that any adoption of such solutions should safeguard the well-being of (potential) employees (Nazareno and Schiff, 2021).

Another important implication of our study is the need to specify the employee's profile that will oppose the recommendation of the IDSS and alert the company if s/he identifies irregularities in how the system works. Our research does not show who is prone to accept the recommendations of the IDSS, but such knowledge is key to recruiting reliable "IDSS's supervisors." On the one hand our results show that HR professionals using clinical judgements may not always be accurate. On the other, we show that an IDSS algorithm used

in statistical judgement, when biased, may also bring negative consequences. It is crucial for HR professionals to try to avoid such influences (of their own misjudgement or biased algorithms). Thus, we support the appeal of other researchers to define the critical competencies of employees of the industry 4.0 (Saniuk *et al.*, 2021) and supplement the previous postulates with such attitudes, which we call "algorithmic nonconformism," i.e. the tendency to oppose IDSS suggestions when they raise ethical doubts of the human supervisor. Therefore, to extend the currently proposed competencies for industry 4.0 (Shet and Pereira, 2021), we shed light on the critical challenge for HRM—discovering who fits? The profile of an algorithmic nonconformist and understanding whether it is possible to teach people such an attitude.

5.1 Limitations

Our research is not free from limitations and, thus, the results should be interpreted accordingly. First, the sample was relatively small and the distribution to experimental groups was unequal. While we controlled for the potential bias, further studies could replicate the experiment on a bigger scale. Furthermore, our subjects were homogenous in terms of their background and experience level. Future studies should examine the effects discovered in this study among professionals of various experience levels and various cultures. As individuals of various cultures perceive technology differently, we may expect possible variations of the results. Finally, due to Covid-19 sanitary restrictions, we had to adjust our study to online settings. Future research should investigate a similar problem in fully controlled lab conditions and, as an additional step, in a live experiment, considering the ethical considerations we mentioned earlier in the discussion.

6. Conclusion

In this paper, we investigate individuals opposing biased algorithm recommendations regarding disciplinary actions in an organisation. We conducted an experiment in which we tested 76 subjects over a set of 5 scenarios in which a biased algorithm gave strict recommendations regarding disciplinary actions for employees violating the disciplinary work code. Thus, we showed that employees follow IDSS recommendations, even if these may be harmful to their colleagues. Our results suggest that humans are harsher in evaluating colleagues when an algorithm suggests more strict disciplinary actions, making their decisions faster. Our results contribute to the HR management and career development literature and the ongoing debate on applying AI solutions to HR problems.

References

Ægisdóttir, S., White, M.J., Spengler, P.M., Maugherman, A.S., Anderson, L.A., Cook, R.S., Nichols, C.N., Lampropoulos, G.K., Walker, B.S., Cohen, G. and Rush, J.D. (2006), "The meta-analysis of clinical judgment project: fifty-six years of accumulated research on clinical versus statistical prediction", *The Counseling Psychologist*, SAGE Publications, Vol. 34 No. 3, pp. 341-382.

Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y. and Chen, H. (2021), "Artificial intelligence in sustainable energy industry: status quo, challenges and opportunities", *Journal of Cleaner Production*, Vol. 289, 125834.

Akerkar, R. (2018), *Artificial Intelligence for Business*, 1st ed. 2019, Springer, Cham.

Al Mansoori, S., Salloum, S. and Shaalan, K. (2021), "The impact of artificial intelligence and information technologies on the efficiency of knowledge management at modern organizations: a systematic review", *Studies in Systems, Decision and Control*, pp. 163-182.

Albrecht, T., Rausch, T.M. and Derra, N.D. (2021), "Call me maybe: methods and practical implementation of artificial intelligence in call center arrivals' forecasting", *Journal of Business Research*, Vol. 123, pp. 267-278.

Allal-Chérif, O., Yela Aránega, A. and Castaño Sánchez, R. (2021), "Intelligent recruitment: how to identify, select and retain talents from around the world using artificial intelligence", *Technological Forecasting and Social Change*, Vol. 169, 120822.

Ananny, M. and Crawford, K. (2018), "Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability", *New Media & Society*, SAGE Publications Sage UK: London, Vol. 20 No. 3, pp. 973-989.

Appen (2021), "The state of AI and machine learning, AI industry accelerates rapidly despite pandemic, driven by data partnerships and increasing budgets".

Asch, S.E. (1955), "Opinions and social pressure", *Scientific American*, Scientific American, a division of Nature America, Vol. 193 No. 5, pp. 31-35.

Balasubramanian, N., Ye, Y. and Xu, M. (2022), "Substituting human decision-making with machine learning: implications for organizational learning", *Academy of Management Review*, Academy of Management, Vol. 47 No. 3, pp. 448-465.

Bigman, Y.E. and Gray, K. (2018), "People are averse to machines making moral decisions", *Cognition*, Vol. 181, pp. 21-34.

Black, J.S. and van Esch, P. (2021), "AI-enabled recruiting in the war for talent", *Business Horizons*, Vol. 64 No. 4, pp. 513-524.

Böckle, M., Novak, J. and Bick, M. (2020), "Exploring gamified persuasive system design for energy saving", *Journal of Enterprise Information Management*, Emerald Publishing, Vol. 33 No. 6, pp. 1337-1356.

Bond, R. and Smith, P.B. (1996), "Culture and conformity: a meta-analysis of studies using Asch's (1952b, 1956) line judgment task", *Psychological Bulletin*, American Psychological Association, Vol. 119 No. 1, pp. 111-137.

Cossins, D. (2018), "Discriminating algorithms: 5 times AI showed prejudice", New Scientist, available at: https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/ (accessed 2 September 2021).

Davenport, T. and Kirby, J. (2016), "Just how smart are smart machines", *MIT Sloan Management Review*, available at: https://www.semanticscholar.org/paper/Just-How-Smart-Are-Smart-Machines-Davenport-Kirby/a9d80b09f21d9d0306766d2c3ba2ce49b4b2b95b (accessed 7 January 2022).

Dawes, R., Faust, D. and Meehl, P. (1989), "Clinical versus actuarial judgment", *Science*, Vol. 243 No. 4899, pp. 1668-1674.

Dourish, P. (2004), *Where the Action is: The Foundations of Embodied Interaction*, New ed., The MIT Press, Cambridge, MA.

Evans, L. and Kitchin, R. (2018), "A smart place to work? Big data systems, labour, control and modern retail stores", *New Technology, Work and Employment*, Vol. 33 No. 1, pp. 44-57.

Fogg, B.J. (2003), *Persuasive Technology. Using Computers to Change What We Think and Do*, Morgan Kaufmann, San Francisco.

Frey, C.B. and Osborne, M.A. (2017), "The future of employment: how susceptible are jobs to computerisation?", *Technological Forecasting and Social Change*, Vol. 114, pp. 254-280.

Galaz, V., Centeno, M.A., Callahan, P.W., Causevic, A., Patterson, T., Brass, I., Baum, S., Farber, D., Fischer, J., Garcia, D., McPhearson, T., Jimenez, D., King, B., Larcey, P. and Levy, K. (2021), "Artificial intelligence, systemic risks and sustainability", *Technology in Society*, Vol. 67, 101741.

Gruetzemacher, R., Paradice, D. and Lee, K.B. (2020), "Forecasting extreme labor displacement: a survey of AI practitioners", *Technological Forecasting and Social Change*, Vol. 161, 120323.

Grundner, L. and Neuhofer, B. (2021), "The bright and dark sides of artificial intelligence: a futures perspective on tourist destination experiences", *Journal of Destination Marketing and Management*, Vol. 19, 100511.

Hancock, P.A. (2017), "Imposing limits on autonomous systems", *Ergonomics*, Vol. 60 No. 2, pp. 284-291.

Hertz, N. and Wiese, E. (2018), "Under pressure: examining social conformity with computer and robot groups", *Human Factors*, Vol. 60 No. 8, pp. 1207-1218.

Huang, T.-L. and Hsu Liu, F. (2014), "Formation of augmented-reality interactive technology's persuasive effects from the perspective of experiential value", *Internet Research*, Emerald Group Publishing, Vol. 24 No. 1, pp. 82-109.

Humerick, J.D. (2019), "Reprogramming fairness: affirmative action in algorithmic criminal sentencing", *HRLR Online*, Vol. 4, p. 213.

Jia, Q., Guo, Y., Li, R., Li, Y. and Chen, Y. (2018), "A conceptual artificial intelligence application framework in human resource management", *ICEB 2018 Proceedings*, Guilin, China, available at: https://aisel.aisnet.org/iceb2018/91.

Keding, C. and Meissner, P. (2021), "Managerial overreliance on AI-augmented decision-making processes: how the use of AI-based advisory systems shapes choice behavior in R&D investment decisions", *Technological Forecasting and Social Change*, Vol. 171, 120970.

Kelley, J.F. (1983), "An empirical methodology for writing user-friendly natural language computer applications", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 193-196.

Kellogg, K.C., Valentine, M.A. and Christin, A. (2020), "Algorithms at work: the new contested terrain of control", *Academy of Management Annals*, Academy of Management, Vol. 14 No. 1, pp. 366-410.

Kirchmer, M. (2017), "Robotic process automation – pragmatic solution or dangerous illusion?", *BTOES Insights (Business Transformation and Operational Excellence Summit Insights)*.

Klien, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R. and Feltovich, P.J. (2004), "Ten challenges for making automation a 'team player' in joint human-agent activity", *IEEE Intelligent Systems*, presented at the IEEE Intelligent Systems, Vol. 19 No. 6, pp. 91-95.

Lankton, N.K., McKnight, D.H. and Tripp, J. (2015), "Technology, humanness, and trust: rethinking trust in technology", *Journal of the Association for Information Systems*, Vol. 16 No. 10, doi: 10.17705/1jais.00411.

Lecher, C. (2019), "How Amazon automatically tracks and fires warehouse workers for 'productivity'", The Verge, 25 April, available at: https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations (accessed 7 January 2022).

Li, C.-Y. (2013), "Persuasive messages on information system acceptance: a theoretical extension of elaboration likelihood model and social influence theory", *Computers in Human Behavior*, Vol. 29 No. 1, pp. 264-275.

Makridis, C.A. and Han, J.H. (2021), "Future of work and employee empowerment and satisfaction: evidence from a decade of technological change", *Technological Forecasting and Social Change*, Vol. 173, 121162.

Malik, A., Budhwar, P., Patel, C. and Srikanth, N.R. (2022), "May the bots be with you! Delivering HR cost-effectiveness and individualised employee experiences in an MNE", *The International Journal of Human Resource Management*, Vol. 33 No. 6, pp. 1148-1178.

Mazari Abdessameud, O., Van Utterbeeck, F. and Guerry, M.-A. (2021), "Military human resource planning through flow network modeling", *Engineering Management Journal*, Vol. 34 No. 2, pp. 302-313.

Meehl, P.E. (1957), "When shall we use our heads instead of the formula?", *Journal of Counseling Psychology*, Wm. C. Brown Co., Vol. 4 No. 4, p. 268.

Meijerink, J., Boons, M., Keegan, A. and Marler, J. (2021), "Algorithmic human resource management: synthesizing developments and cross-disciplinary insights on digital HRM", *The International Journal of Human Resource Management*, Routledge, Vol. 32 No. 12, pp. 2545-2562.

Möhlmann, M., Zalmanson, L., Henfridsson, O. and Gregory, R.W. (2021), "Algorithmic management of work on online labor platforms: when matching meets control", *MIS Quarterly: Management Information Systems*, pp. 1-54.

Mora, M., Forgionne, G., Cervantes, F., Garrido, L., Gupta, J.N.D. and Gelman, O. (2005), "Toward a comprehensive framework for the design and evaluation of intelligent decision-making support systems (i-DMSS)", *Journal of Decision Systems*, Taylor & Francis, Vol. 14 No. 3, pp. 321-344.

Nass, C. and Moon, Y. (2000), "Machines and mindlessness: social responses to computers", *Journal of Social Issues*, Vol. 56 No. 1, pp. 81-103.

Nazareno, L. and Schiff, D.S. (2021), "The impact of automation and artificial intelligence on worker well-being", *Technology in Society*, Vol. 67, 101679.

Nissen, M.E. and Sengupta, K. (2006), "Incorporating software agents into supply chains: experimental investigation with a procurement task", *MIS Quarterly*, Management Information Systems Research Center, University of Minnesota, Vol. 30 No. 1, pp. 145-166.

Novichkov, A.V., Puzynya, T.A., Grishina, T.V., Fursova, S.D. and Buley, N.V. (2021), "The impact of artificial intelligence on retraining", in Bogoviz, A.V., Suglobov, A.E., Maloletko, A.N., Kaurova, O.V. and Lobova, S.V. (Eds), *Frontier Information Technology and Systems Research in Cooperative Economics. Studies in Systems, Decision and Control*, Springe, Cham, Vol. 316, pp. 469-476.

Obedkov, E. (2021), "Xsolla fires 150 employees using big data and AI analysis, CEO's letter causes controversy", *Game World Observer*, 4 August, available at: https://gameworldobserver.com/2021/08/04/xsolla-fires-150-employees-using-big-data-and-ai-analysis-ceos-letter-causes-controversy/ (accessed 7 January 2022).

Oleson, J.C., VanBenschoten, S.W., Robinson, C.R. and Lowenkamp, C.T. (2011), "Training to see risk: measuring the accuracy of clinical and actuarial risk assessments among federal probation officers", *Federal Probation*, Administrative Office of the United States Courts, US, Vol. 75 No. 2, pp. 52-56.

Orlikowski, W.J. and Scott, S.V. (2014), "What happens when evaluation goes online? Exploring apparatuses of valuation in the travel sector", *Organization Science*, INFORMS, Vol. 25 No. 3, pp. 868-891.

Orriols-Puig, A., Martínez-López, F.J., Casillas, J. and Lee, N. (2013), "Unsupervised KDD to creatively support managers' decision making with fuzzy association rules: a distribution channel application", *Industrial Marketing Management*, Vol. 4 No. 42, pp. 532-543.

Ostheimer, J., Chowdhury, S. and Iqbal, S. (2021), "An alliance of humans and machines for machine learning: hybrid intelligent systems and their design principles", *Technology in Society*, Vol. 66, 101647.

Pan, Y., Froese, F., Liu, N., Hu, Y. and Ye, M. (2022), "The adoption of artificial intelligence in employee recruitment: the influence of contextual factors", *The International Journal of Human Resource Management*, Vol. 33 No. 6, pp. 1125-1147.

Peng, C.-H., Lurie, N.H. and Slaughter, S.A. (2019), "Using technology to persuade: visual representation technologies and consensus seeking in virtual teams", *Information Systems Research*, INFORMS, Vol. 30 No. 3, pp. 948-962.

Phillips-Wren, G., Mora, M., Forgionne, G.A. and Gupta, J.N.D. (2009), "An integrative evaluation framework for intelligent decision support systems", *European Journal of Operational Research*, Vol. 195 No. 3, pp. 642-652.

Pocock, S.J. (1979), "Allocation of patients to treatment in clinical trials", *Biometrics*, Vol. 35 No. 1, pp. 183-197.

Quintero, A., Konaré, D. and Pierre, S. (2005), "Prototyping an intelligent decision support system for improving urban infrastructures management", *European Journal of Operational Research*, Vol. 162 No. 3, pp. 654-672.

Rosander, M. and Eriksson, O. (2012), "Conformity on the Internet – the role of task difficulty and gender differences", *Computers in Human Behavior*, Vol. 28 No. 5, pp. 1587-1595.

Salomons, N., Sebo, S.S., Qin, M. and Scassellati, B. (2021), "A minority of one against a majority of robots: robots cause normative and informational conformity", *ACM Transactions on Human-Robot Interaction*, Vol. 10 No. 2, pp. 15:1-15:22.

Saniuk, S., Caganova, D. and Saniuk, A. (2021), "Knowledge and skills of industrial employees and managerial staff for the industry 4.0 implementation", *Mobile Networks and Applications* (in press).

Schaefer, K.E., Chen, J.Y.C., Szalma, J.L. and Hancock, P.A. (2016), "A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems", *Human Factors*, SAGE Publications, Vol. 58 No. 3, pp. 377-400.

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I. and Lowry, P.B. (2020), "Collaborating with technology-based autonomous agents: issues and research opportunities", *Internet Research*, Emerald Publishing, Vol. 30 No. 1, pp. 1-18.

Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002), *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton, Mifflin and Company, Boston, MA, pp. xxi, 623.

Shet, S.V. and Pereira, V. (2021), "Proposed managerial competencies for Industry 4.0 – implications for social sustainability", *Technological Forecasting and Social Change*, Vol. 173, 121080.

Shin, Y. and Kim, J. (2018), "Data-centered persuasion: nudging user's prosocial behavior and designing social innovation", *Computers in Human Behavior*, Vol. 80, pp. 168-178.

Song, Y. and Wu, R. (2021), "Analysing human-computer interaction behaviour in human resource management system based on artificial intelligence technology", *Knowledge Management Research and Practice* (in press).

Tambe, P., Cappelli, P. and Yakubovich, V. (2019), "Artificial intelligence in human resources management: challenges and a path forward", *California Management Review*, SAGE Publications, Vol. 61 No. 4, pp. 15-42.

Trocin, C., Hovland, I.V., Mikalef, P. and Dremel, C. (2021), "How Artificial Intelligence affords digital innovation: a cross-case analysis of Scandinavian companies", *Technological Forecasting and Social Change*, Vol. 173, 121081.

Upchurch, M. (2018), "Robots and AI at work: the prospects for singularity", *New Technology, Work and Employment*, Vol. 33 No. 3, pp. 205-218.

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A. and Trichina, E. (2022), "Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review", *The International Journal of Human Resource Management*, Vol. 33 No. 6, pp. 1237-1266.

Zarsky, T. (2016), "The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making", *Science, Technology and Human Values*, SAGE Publications, Vol. 41 No. 1, pp. 118-132.

Corresponding author
Marcin Lukasz Bartosiak can be contacted at: marcin.bartosiak@unipv.it