Received: 8 April 2020 | Revised: 13 December 2021 | Accepted: 20 December 2021

DOI: 10.1111/1748-8583.12433

Human Resource Management Journal WILEY

REVIEW ARTICLE

Can HR adapt to the paradoxes of artificial intelligence?

Andy Charlwood[1] | **Nigel Guenole**[2]

[1]University of Leeds, Leeds, UK

[2]Goldsmiths, University of London, London, UK

Correspondence
Andy Charlwood, University of Leeds, Leeds, UK.
Email: a.charlwood@leeds.ac.uk

Abstract

Artificial intelligence (AI) is widely heralded as a new and revolutionary technology that will transform the world of work. While the impact of AI on human resource (HR) and people management is difficult to predict, the article considers potential scenarios for how AI will affect our field. We argue that although popular accounts of AI stress the risks of bias and unfairness, these problems are eminently solvable. However, the way that the AI industry is currently constituted and wider trends in the use of technology for organising work mean that there is a significant risk that AI use will degrade the quality of work. Viewing different scenarios through a paradox lens, we argue that both positive and negative visions of the future are likely to coexist. The HR profession has a degree of agency to shape the future if it chooses to use it; HR professionals need to develop the skills to ensure that ethics and fairness are at the centre of AI development for HR and people management.

KEYWORDS
artificial intelligence, human resource management

1 | INTRODUCTION

Artificial intelligence (AI) is widely heralded as a new and revolutionary general purpose technology that will transform the world of work (Byrnjolfsson & Macafee, 2014; Agrawal et al., 2018). While there are a number of challenges

Abbreviations: AI, artificial intelligence; GDPR, General Data Protection Regulation; HR, human resources; HRM, human resource management; IO, industrial/organizational; ML, machine learning.

Practitioner notes

- The use of artificial intelligence (AI) for human resources (HR) and people management is currently in its infancy
- It is possible to conceive of optimistic and pessimistic accounts of how AI might affect HR and people management. A paradox lens suggests both will likely coexist in our immediate future
- Without regulation, existing approaches to people management could lead to AI that dramatically reduces worker autonomy and ramps up effort and stress
- The ethical values and practical insights of the HR profession are important if this 'bad AI' is to be contained
- An ethical approach to AI for HR involves the full involvement of workers and stakeholders in the design and deployment of AI systems

in developing AI for human resource management (Tambe et al., 2019), a recent industry study found 300 plus human resources (HR) technology start-ups developing AI tools and products for HR or people management, with around 60 of these companies 'gaining traction' in terms of customers and venture capital funding (Bailie & Butler, 2018). Recently, eightfold, an AI based talent intelligence platform that helps attract, develop and retain top talent, has raised $220 million and is now valued at over $2 billion (Singh, 2021). Phenom, an HR technology firm that automates job tasks and improves the job search experience, was reported by Forbes as having 'quietly become a billion dollar unicorn' (Kelly, 2021). Accenture has made a strategic investment in the London based start-up Beamery, which offers its own operating system for recruitment, and has a reported valuation of $800 million (Lunden, 2021). Large global technology firms have also started to make AI central to their people management systems and processes (van den Broek et al., 2021); IBM reported that in a single year its cost savings due to implementing artificial intelligence in human resources exceeded $100 million (Guenole & Feinzig, 2019).

Although deployment of AI systems in business is as yet limited (Benbya et al., 2020), examples like those cited above suggest an AI driven future is fast approaching. In this context there are widespread popular concerns about whether the use of AI in HR will lead to a dystopian work future, where AIs as managers degrade the quality of work (Dzieza, 2020). These concerns reflect broader insider critiques of the technology industry, which argue that AI developers are overly focussed on technical and commercial priorities and neglect the ethical and societal impact of their work (Birhane, 2021; Crawford, 2021; Whittaker, 2021).

The contribution of this article is to bring together contrasting accounts of how AI might be deployed for HR and people management to address the question of how AI could affect our field. It is possible to imagine a future in which the deployment of AI for HR and people management leads to large gains in fairness and efficiency. However, AI could also usher in an increasingly dystopian future of widespread unfairness and intensified managerial control. Indeed, a paradox lens suggests that both imaginaries will co-exist alongside each other (Collings et al., 2021; Smith & Lewis, 2011). The future that emerges will depend on choices we all make. The HR profession must be engaged in shaping these choices. Without HR involvement, HR work risks being de-skilled and replaced by black-box algorithms (Callen, 2021), locking organisations into an approach to people management and patterns of work that we hoped we had progressed beyond (Raisch & Krakowski, 2021).

The article begins by defining and explaining what AI is, both generally and as it pertains to the field of HR. Our focus is on how AI will affect the HR profession, its activities and the activity of people management more broadly. This means that we do not explicitly address some of the broader debates about how AI will affect the quantity and types of jobs (Frey & Osborne, 2017; Susskind & Susskind, 2015). We then consider alternative scenarios for how AI might be used. We argue that the risks that the realities of AI in HR will be increasingly dystopian are real, but there are solutions to many of the emerging problems if we have the will to work for them. We finish by considering how

those of us who study and practice HR can take action to move towards the more desirable future, considering what the HR profession can do and setting out a future research agenda.

2 | WHAT IS AI?

AI is typically defined as the use of digital technology to create systems capable of autonomously performing tasks commonly thought to require human intelligence (Office for AI, 2019). In contrast to popular representations of artificial general intelligence in science fiction, recent advances in AI have occurred in the field of machine learning (ML), a sub-set of AI where digital systems autonomously improve their performance at undertaking a specific task or tasks over-time as the system learns through experience (Office for AI, 2019). Over the last decade there have been significant advances in the use of ML AI in text analysis, speech recognition and comprehension and image recognition.

Conceptually, ML AI is often described as making predictions (Agrawal et al., 2018) or classification (Birhane, 2021; Crawford, 2021; Whittaker, 2021). An AI first identifies a classification schema and then predicts which categories within the schema new cases fall into. It takes a lot of effort and data to get AI to successfully make the first prediction but once trained, the cost of subsequent predictions should be very low (Agrawal et al., 2018). A complicating factor are 'edge-cases'; new scenarios with features that the AI has not encountered before so cannot initially classify, for example, a candidate with a speech impediment in an automated video interview. The underlying issue here is the quality of the training data; is there a sufficiently representative set of data labelled sufficiently well for ML AI to work (Lebovitz et al., 2021)?

Over 100 papers have been published on technical aspects of applying ML AI to HR and people management (Jatobá et al., 2019; Strohmeier & Piazza, 2013) but there is limited evidence on the use and consequences of ML AI to our field in practice [van den Broek et al. (2021) are a recent exception], with most published empirical research focussed on investigating responses to hypothetical scenarios (Langer & Landers, 2021). Because AI is a new general purpose technology many of the eventual uses that it will be put have yet to be conceived (Byrnjolfsson & Macafee, 2014); therefore AI is likely to have unforeseen consequences. Any attempt to assess its expected impacts on our field must proceed with (and be read with) caution. Nevertheless, it is important to study the phenomenon now, because this is the moment at which it is, perhaps, possible to shape and improve on the direction of travel (Bailey & Barley, 2020; Benbya et al., 2020; Glickson & Woolley, 2020; Whittaker, 2021).

3 | AN OPTIMISTIC VISION OF THE FUTURE

Central to an optimistic vision of the future is the idea that ML AI can dramatically improve the efficiency and fairness of how people are managed (Chamorro-Premuzic et al., 2019; Guenole & Feinzig, 2019). We will explore how this might happen through investigation of AI use cases currently being developed. In many ways, the focus of these use cases represents a continuation of existing trends in HR technology. What marks out AI as new and different is the scale, accuracy and efficiency of the cognitive tasks that it can undertake.

Among the most common use cases is AI that resolves the twin difficulties businesses face in identifying applicant pools of people with the knowledge, skills, abilities and attitudes that the business needs and in making cost-effective decisions about who to select from available applicant pools. While existing application management systems can partially automate selection by filtering out applications without key words relevant to the post, the promise of AI is that new systems will be able to seek out and advertise posts to new applicant pools (doing the job currently done by recruitment consultants more cheaply and at greater scale) and then automate early stages of selection by using a combination of more sophisticated algorithmic analysis of CVs and applications, gamified applicant tests and robot interviews that select candidates with job relevant traits and skills that mirror those of the businesses existing top employees. The use of ML AI will mean that organisations can recruit and select highly efficiently from larger

applicant pools without selection decision-making being biased by recruiter heuristics (e.g. educational background) and biases (e.g. gender and ethnicity) that have no relevance for job performance.

If this use case becomes established, the same principles and can be applied to decision making about workforce planning and talent management within organisations. ML AIs can use a range of digital data from employee software use, communications, manufacturing and service delivery censors, and audio and video image feeds to judge employee performance more accurately and fairly than human managers are able to do, and then use this information to make or recommend decisions such as which staff to employ, where and when, who to promote into which role, and which staff to award pay rises to in order to optimise motivation and retention (Angrave et al., 2016; Charlwood, 2021; Guenole & Feinzig, 2019; Kellogg et al., 2020).

AI could also dramatically improve the efficiency and quality of HR operations more generally through chatbot use (Strohmeier & Piazza, 2015), which is now a common application of AI technology in HR. HR chatbots allow humans (e.g., workers, prospective applicants) to interact with virtual agents in natural language. The chatbot can respond instantly to user prompts in any language, at any time of day and on any day of the year. Chatbots can provide information and in certain cases can even complete tasks (e.g., actioning request for an employment reference, or adding calendar notifications for events such as planned leave or travel, booking places on training courses), allowing human experts to focus on more complicated queries where human expertise adds value.

4 | A PESSIMISTIC CASE

Our pessimistic case focuses on possible consequences of the sorts of use cases set out above for workers and societies. The central problem is that firms tend to introduce new technologies in ways that reduce worker autonomy, wages, and job security.

5 | MANAGEMENT, LABOUR AND TECHNOLOGY

Labour process theory posits that firms are compelled by the logic of profit seeking to constantly change production and service delivery to bring down labour costs, and that the indeterminacy of labour creates a strong tendency to make greater use of surveillance and monitoring technologies in pursuit of this goal (Thompson, 2010: p. 10; Kellogg et al., 2020). A range of evidence supports these propositions. Procurement of previous technological innovations has depended on technology advocates framing the case for the technology around increased efficiencies and lower labour costs (Bailey & Barley, 2020; Thomas, 1994). Globally, the decline in the share of income going to labour since 1970 has been driven by increasing investment in information technology assets (O'Mahoney et al., 2020). Declining worker power and technology use have resulted in workers working harder with reduced autonomy and control over how they work (Green et al., 2021; Williams et al., 2020). The key question then is how will this general tendency to adopt technologies that improves efficiency while degrading work interact with the specific technology of ML AI?

6 | THE TECHNOLOGY INDUSTRY AND COMMERCIALISATION OF AI

Automation and augmentation represent different ideological paradigms for AI design (Markoff, 2016). The logic of labour process theory is that potential purchasers of AI are likely to favour automation, so AI developers will likely gravitate to this approach (Kochan, 2021). The ethical and societal consequences of a widespread automation approach are likely to be downplayed because developers take only a narrow view of ethics; commercial considerations come first (Crawford, 2021; Simonite, 2021). Broader critiques of the technology industry suggest a disregard for ethics and societal impact is hard-wired into it (Birhane, 2021; Crawford, 2021; Whittaker, 2021; Zuboff, 2019).

The issue here is not that problems of biased and unfair AIs cannot be solved. It is that developers of AI have no interest in trying to solve the ethical and societal problems that the deployment of AI might cause (Birhane, 2021; Crawford, 2021). Critics may question the internal logics that drive AI classification schema (Crawford, 2021) or argue that AI use cases in the field of HR are nothing more than snake-oil because the internal logics of classification systems are fundamentally flawed but as yet these critiques do not appear to have dented the intellectual confidence of AI advocates in offering 'snake-oil' products (Narayan, 2019).

7 | THE EARLY REALITIES OF ML AND AI IN PEOPLE MANAGEMENT

The retail and distribution sectors offer a taste of the new world of work that is being ushered in by AI ML. The relative simplicity of tasks has allowed for substantial use of algorithms to classify and predict the most efficient way of working, extracting maximum effort from workers while only offering low paid and insecure jobs in a form of digital Taylorism (Moore, 2019). Most notoriously, accounts of Amazon warehouses are replete with examples of workers sustaining serious injuries and working to the point of absolute physical exhaustion as a result of the pace and intensity of work (Bloodworth, 2018; O'Connor, 2021). Amazon has attracted further opprobrium for automating workers dismissals (Bort, 2019). In the retail sector scheduling algorithms seek to minimise retailer labour use while also minimising the numbers of workers who work enough hours to qualify for the enhanced benefits, resulting in instability of hours and income (Schulte, 2020; Ton, 2012). Algorithmic management tools with similar effects are being developed and deployed in the nascent 'gig work' sector of technologically mediated, hyper flexible employment (Duggan et al., 2020; Möhlmann et al., 2021).

To date, the use of algorithmic management has been confined by the limits of technology to controlling relatively simple tasks where relatively structured data is generated automatically through sensors. The ability of AI to recognise images and undertake more complex pattern recognition tasks is likely to increase monitoring and reduce worker autonomy in new sectors of employment. Bars and restaurants are starting to use their security cameras to feed AIs that monitor and manage the performance of serving staff (Matsakis, 2019). Hospitals are starting to use algorithm-based systems to prioritise and allocate nursing tasks. Amazon has recently started to roll out use of AI linked cameras to monitor the activities of its delivery drivers (Sonnemaker, 2021). The broad concern here is that AI facilitates a world of constant surveillance at work.

Arguably, these problematic examples arise from aspects of the ideology of AI ML developers that means they favour automation of decision-making over augmentation (Markoff, 2016) because they believe that using AI ML will result in better outcomes than traditional decision making by domain experts. The problems of this approach are carefully described in a recent ethnographic study of the development of an ML tool for hiring graduate recruits in a multinational company (van den Broek et al., 2021). Developers had little domain knowledge, and actively sought to exclude domain knowledge from the development of the tool, favouring a purely data-driven approach. Domain experts in the company commissioning the tool successfully pushed back, arguing that without domain knowledge it was impossible to determine the employee performance characteristics the ML tool was trying to predict and that hiring decisions would be taken on the basis of predictors that had no causal impact on employee performance. Dangerous cases of AI use are likely to occur where domain experts are excluded from the design and development of AI tools.

8 | IS A BETTER FUTURE POSSIBLE?

The optimistic and pessimistic accounts set out above are not mutually exclusive. A paradox lens suggests that the development of AI will be shaped by contradictory demands which will result in aspects of both co-existing (Smith & Lewis, 2011). The key question then is what can be done to contain the negative and promote the positive? In

answering this question, we stake out two claims. First, problems of AI bias are eminently solvable. Second, while the threat of AI snake-oil is real, it is possible to develop fair, ethical and efficient AI ML if design is informed by domain knowledge.

9 | MANAGING THE BIAS CHALLENGE

In responding to Google's dismissal of her colleague Timit Gebru, Margaret Mitchell (who was also subsequently dismissed, Simonite, 2021) made an insightful comment about the nature of bias in artificial intelligence: 'When diving deep into operationalising the ethical development of artificial intelligence, one immediately runs into the "fractal problem". This may also be called the "infinite onion" problem. That is, each problem within development that you pinpoint expands into a vast universe of new complex problems. It can be hard to make any measurable progress as you run in circles among different competing issues, but one of the paths forward is to pause at a specific point and detail what you see there.' Here we try to pause, look around, and describe what we see with reference to the bias debate around AI in HR systems.

10 | DEFINING BIAS

In the context of AI in HR, adverse impact is a technical term referring to majority and minority group differences in the employment related opportunities, for example, hiring, promotion and termination, that are distributed as a result of using an AI system. The differences in the opportunities distributed across groups may result from real differences between groups, or they may be due to AI models measuring differently or predicting differently for different groups. Until recently, the computer science community has not had a method for differentiating these two possibilities, although recent work has started link the social science and computer science fields on this issue (Hutchinson & Mitchell, 2019).

In any event, improving models so that they do function in the same way for all groups does not always sufficiently reduce adverse impact to the point of no concern. Under North American Civil Rights law, if any differences are sufficiently large to violate the 4/5 rule as evidenced by a statistical test (i.e. if the selection rate for a group is less than 80% of the group with the highest selection rate), then irrespective of whether they are due to real differences between groups or due to the AI system measuring or predicting differently, the AI system is producing adverse impact (EEOC, 1979; Scherer, 2017). In the computer science literature, the terms adverse impact, bias, and fairness are used interchangeably to describe this scenario, and indeed, that seems to be the way the media and popular culture understand bias too.

To industrial psychologists, however, this would only be considered adverse impact. It does not demonstrate bias, as it is quite possible to have adverse impact with an AI system that measures and predicts equivalently across groups (i.e., the AI bases decisions on real differences between groups), that is, you can have a system that is unbiased but still produces adverse impact. This distinction is of fundamental importance, because removing adverse impact caused by true differences from an AI system typically leads to lower predictive accuracy of future job performance (Pyburn et al., 2008). Social scientists call this the diversity-validity dilemma. In short, some organisations may be comfortable with different selection rates based on an AI tool, if the decisions are being made on job related criteria, and if the scores from the AI system predicts performance in the same way for all groups. Indeed, this is the way the US courts interpret the situation, which much of the world looks to as the most evolved example of how to regulate adverse impact. Adverse impact does not violate US legislation if the test scores are predictive of performance (albeit organisations do have to show no alternative predictor was available that was as effective and showed lower adverse impact).

Psychologists and statisticians also have a well-developed set of analytical tools for dealing with bias/adverse impact that are primarily content related, they use alternative tests or predict broader performance criteria (for a very detailed review of such methods, see Ployhart & Holtz, 2008). Computer scientist have also developed new methods to tackle adverse impact (Bellamy et al., 2018) that are primarily methodological. They include pre-processing techniques that can transform the predictor set to retain as much information as possible while divorcing the predictor variables of any statistical relationship to protected attributes like race or gender; training constraints that force models to deliver balanced predictions in the model building phase can also be employed; and post-processing methods, for instance, using a coarser scoring system that makes similar scores equal.

11 | MISCONCEPTIONS

A typical objection at this stage, is that the types of data that would be used by ML AI are not good enough for the task of predicting job related outcomes; 'most HR data is bad data' (Buckingham, 2015) because HR data are based on subjective opinions and judgements that are themselves likely to contain biases against minority groups. We think this objection is mistaken. Advanced latent variable models exist today that allow measurement and prediction of latent variables based on inferences about behaviours that can be seen and measured (e.g., questionnaire responses; Bollen, 1989; Kline, 2015) and we can examine whether these work equally well for all groups.

A second objection is that because training data is often based on data regarding white males, models will inevitably be biased in favour of white males (as was the case Amazon's abandoned hiring ML system, described by Bort, 2019). Yet, no well-trained IO psychologist would build a selection system that naively replicates existing top performers, including their gender and ethnicity, and this is not how well-designed AI systems operate either. First, a job analysis would be carried out that identified the knowledge, skills, abilities and attributes (KSAOs) that are required for successful performance. Next, assessments are designed that assess these attributes. It is the difference in scores on these *job-related* attributes that determine candidate ranking in competitive hiring situations. The key point here is that domain knowledge can ensure that AI ML systems do not reproduce the discriminatory behaviours produced by human biases if it is used in the development of AI systems. Job analysis alone will *not* resolve the problem of adverse impact, but it can help resolve the problem of new hires mirroring the performance irrelevant demographic characteristics of past successful candidates. Examples of AI systems where domain knowledge was not applied and which failed as a result do not mean that discriminatory AIs are inevitable.

A final objection is that ML AI models are all black boxes and cannot be examined closely to see why there is adverse impact. But AI systems are only functionally black boxes, in the sense that it is impractical to step through every input to decision a system makes. There are methods for understanding what is happening inside the black box, although the complexities involved make it impractical in many circumstances. Furthermore, model explainability is an active and fast-developing field of research in artificial intelligence (Holzinger et al., 2019) and there are numerous techniques for evaluating variable importance today (e.g., Shap values: Sundararajan & Najmi, 2020) which mean that our ability to explain black box AI models is constantly improving.

Our overall point here is that biased AIs are not inevitable. There are a number of tools and methods that can be used to develop AIs that are largely free of bias, and we can use existing industrial psychology principles to manage adverse impact. The challenge is to ensure that these methods are widely adopted in our field. What can we do to bring this about?

12 | THE ROLE OF THE HR PROFESSION

It would perhaps be tempting for the HR profession to adopt the critiques of Crawford (2021), Birhane (2021), Narayan (2019) and Zuboff (2019) in arguing that AIs cannot accurately classify human behaviour at work and that the use of AI in HR should be resisted because it represents an ideological project to control human behaviour that replaces autonomy with dependency, diminishing our common capacity for flourishing. This would be a natural position for many in the profession to take given widespread scepticism about the quantification of HR (Greasley & Thomas, 2020). However, we think this impulse should be resisted because AI in HR and people management is likely to happen whether we like it or not. Engagement will likely result in better outcomes than unsuccessful resistance, and we have outlined what we believe is a strong case for a positive future for AI in HR.

Domain knowledge and expertise is essential for the development of AI tools that work as intended, are fair and which do not reproduce existing organisational biases (van den Broek et al., 2021; Jacobs & Wallach, 2021). It is also important for senior and experienced professionals to work closely with those who develop and operate automating technologies to ensure that their knowledge and expertise is used to inform use and understanding of the automating algorithms (Callen, 2021). While augmentation and automation are often portrayed as alternative approaches to designing and deploying AI, in practice AI systems are likely to be more effective if the two approaches are combined and successful combination is likely to depend on the incorporation of HR professionals expert knowledge (Raisch & Krakowski, 2021; van den Broek et al., 2021).

Achieving fairness in AI systems is not just a matter of applying domain knowledge alongside appropriate technical fixes. Processes of development and deployment matter too. AI can only be ethical if it is based on consultation with and the involvement of stakeholders (particularly employees and prospective employees) who will be affected by the AI at the design, development and deployment stages (Leslie, 2019). Without HR involvement in negotiating the development of AI with those affected, the risks of unfair and biased AIs that negatively affect workers will likely increase.

Engagement is also needed to secure a future for the HR profession that is worth having. The use of ML AI in HR has the potential to significantly disrupt the routines of HR work in ways that are difficult to predict (Murray et al., 2021). Evidence from other knowledge professions that have been extensively exposed to algorithmic automation suggests that those who use these systems without questioning or understanding the recommendations that the system produces become de-skilled, rapidly losing the professional expertise and commercial acumen that are the basis of their role and status. By contrast, when knowledge workers constantly question and interrogate algorithmic systems and configure social interactions that develop and maintain their expertise they are able to maintain status (Callen, 2021).

If the HR profession is to put its espoused ethical values regarding the importance of human dignity and justice (e.g. AHRI, 2016; CIPD, 2020; CPHR, 2016; SHRM, 2014) into practice and safeguard its own future role and status, individual HR practitioners need to gain the skills to play a full role in developing and implementing AI ethically. This will require upskilling. Scholz (2020) outlines the components of what he calls 'big data literacy': an understanding of the principles of computation, statistics and critical thinking. An understanding of these things confers an understanding of what AI can realistically do, how it does it and what might go wrong. It is important that HR professional recognise they have a degree of personal responsibility for the use of AI in their organisations, instead of speaking the language of responsible and ethical management while evading responsibility for taking action (Grigore et al., 2021).

Human resources professional bodies could also do institutional work (Lawrence & Suddaby, 2006) to promote voluntary standards to guide procurement and development of AI for HR tools and systems (Moore, 2020). Such standards could include a requirement for suppliers to have diverse engineering teams, both because it is not ethically acceptable to exclude representatives of diverse communities from design teams if those communities will be affected by the AI system and because there is evidence to suggest that more diverse teams produce less biased AIs (Cowgill et al., 2020). There could be standards for transparency around the origins of training data used in models and data curation practices to promote the availability of training datasets less likely to result in bias (Jo &

Gebru, 2019), and standards for model explainability, so that it is clear which features are important in producing AI outputs. Use of data sheets that document all aspects of the provenance of models provide one tool for transparency (Hutchinson & Mitchell, 2019). Standards should also include processes for stakeholder consultation and engagement because an ethical approach to AI requires that those affected by the AI should be able to participate in decisions over its development (Berg, 2019; Leslie, 2019; Tambe et al., 2019). The Recruitment and Employment Confederation has recently produced guidance that starts to address these points, but there is scope to go further (REC, 2021). Voluntary standards alone are unlikely to be enough to kerb 'bad AI' in our field. Can statutory regulation provide necessary safeguards?

13 | REGULATING AI FOR HR

In the USA, civil rights legislation provides some protection for workers because employers will be legally vulnerable if they use AI tools that result in adverse impact for minority groups unless they can prove that adverse impact is the result of factors that are predictive of job performance while also protecting worker privacy outside of work (Dattner et al., 2019; Scherer, 2017). However, while this body of law might constrain the development of discriminatory AIs for decision making over hiring, promotions and pay, it does little to constrain AI being used to increase employer dominion over workers while at work through new tools for monitoring and controlling worker activity.

In Europe, the European Commission is looking to build on the protections for workers provided by the General Data Protection Regulation (GDPR) through new regulation on AI, a draft of which was published in April 2021. The regulation posits that the use of AI for hiring, promotion, pay decision making and for management and control of workers is 'high risk' requiring significant safeguards to be put in place if AI is to be used for these purposes. However, critics point to the absence of enforcement mechanisms, with employers left to decide how to manage the risks themselves through adherence to a set of standards and worry that the EU regulation could undermine more stringent national statutes (De Stefano & Aloisi, 2021; Veale & Borgesius, 2021); a number of European states already have laws which state that algorithmic management tools can only be introduced if agreed with worker representatives through co-determination processes (Aloisi & Gramano, 2019).

More broadly, AIs as managers throw up a set of legal challenges that existing employment law in most countries is ill-equipped to deal with. For example, who is legally accountable for decisions made by AI? (Adams-Prassl, 2019). Therefore if we want to avoid 'bad AI' in HR and people management, new laws to deal with the challenges that AI poses will be needed regardless of where we are in the world. Given the ethical values of the AI profession, and the threat to these values posed by unscrupulous employers using AI, logically HR professional bodies should be advocating for such regulation. However, it is also important to recognise the difficulties involved in bringing new regulation about given existing power relationships between technology companies and states (Crawford, 2021; Gebru, 2021; Whittaker, 2021) and between labour and capital within states. Therefore embedding pluralism within HR practice (Dundon & Rafferty, 2018), for example, by recognising the legitimacy of independent worker representation through trade unions so that workers have access to countervailing power resources that can offer protection from exploitation and discrimination by AIs (Moore, 2020), is important too: 'When workers have power, it creates a layer of checks and balances on the tech billionaires whose whim-driven decisions increasingly affect the entire world' (Gebru, 2021).

14 | FUTURE RESEARCH DIRECTIONS

Finally, what role can academic research play? Academic research into AI is particularly needed now because this is the moment at which research can influence how AI use in organisations develops (Bailey & Barley, 2020; Benbya et al., 2020; Glickson & Woolley, 2020; Whittaker, 2021). The relative novelty of AI mean that as yet, there have been

only a few attempts to theorise the role and impact of AI in HR and in the field of management more broadly. Murray et al. (2021) have theorised new forms of agency that AIs and humans may jointly enact. Makarius et al. (2020) have theorised that the employee-AI relationship will depend on the novelty and scope of the AI being deployed and put forward a framework to explain how organisations can deploy AI. Raisch and Krakowski (2021) have critiqued extant thought leadership that sees augmentation and automation as competing design approaches, and posit that automation and augmentation approaches are interdependent in practice. Teodorescu et al. (2021) have developed a typology augmentation approaches and consider the research questions that arise from this typology. In our specific field of HR, Prikshat et al. (2021) put forward a structural-functionalist theoretical model to predict the conditions that will lead AI adoption and the likely consequences of different antecedents and processes of adoption. These articles all contain interesting ideas and insights, but we tend to agree with Von Krogh (2018) that research into AI in our field is at the pre-theoretic stage. What is needed is qualitative phenomenological research that can provide the basis for novel theoretical insights.

Bailey and Barley (2020) echo this with a call for ethnographic field studies that unpick how AI is being used in practice. How do they reconfigure divisions of labour and social relations within organisations? This type of research would be particularly valuable for understanding how AI is affecting the HR profession; how are the tasks of HR practitioners changing, is AI up-skilling or de-skilling them? How much agency do they have to moderate the impact of AI on their work? How does augmentation work in practice, that is, how do HR practitioners and managers use the outputs and recommendations of AI systems in their activities and decision-making? Questions that the evidence and analysis of van den Broek et al. (2021) have started to address, illustrating the rich potential for this sort of research.

Bailey and Barley also call for research that extends beyond the social construction of AI design and use. They argue for research that considers the ideologies and power dynamics of those who develop and commission AI for the purposes of managing and organising. What are designers trying to achieve and why? What factors, ideological and institutional, shape the decisions of those procuring (or not) AI systems? In our field, this points to the need to study the growing business eco-system of AI for HR start-ups identified by Bailie and Butler (2018). What areas of HR activity are their products focussed on? What are their motivations and intentions in developing their products? Who are their customers and what are their aims and objectives in procuring AI for HR systems? Research to uncover this new institutional field might include mapping surveys, ethnographic studies of HR tech start-ups, and ethnographic engagement with industry experts to identify broader trends and issues.

Finally, it is also important to study the consequences of AI on workers and by extension society, including the professional and societal behaviours, norms and institutions that arise from the use of AI in our field (Bailey & Barley, 2020; Langer & Landers, 2021). Audit studies of organisations using AI and ML tools for recruitment and selection can reveal the extent to which the use of these tools reproduces, neutralises or intensifies the forms of discrimination that audit studies of traditional selection methods have revealed. How does the use of AI in HR and people management affect worker wellbeing and change the norms and the values they attach to work and employment? How do trade unions and NGOs respond and engage with firms introducing AI for people management tasks and activities? How do those doing the work of developing ML AI in our field respond to these societal–institutional forces shaping their activities? Do they seek to establish their own professional and ethical norms and standards?

ACKNOWLEDGEMENTS

We are grateful to Geoff Wood, Anthony McDonnell and three anonymous referees for helpful comments on previous draughts of the article and to Martin Edwards for the initial encouragement to develop it. As part of the Digital Futures at Work Research Centre (Digit), Charlwood's work was supported by the UK Economic and Social Research Council (grant number ES/S012532/1), which is gratefully acknowledged. Guenole received no external funding.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed.

ORCID

Andy Charlwood https://orcid.org/0000-0002-5444-194X

REFERENCES

Adams-Prassl, J. (2019). What if your boss was an algorithm? The rise of artificial intelligence at work. *Comparative Labor Law & Policy Journal*, *41*(1), 123–146.

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Review Press.

AHRI. (2016). Code of ethics and professional conduct. Retrieved December 18, 2020, from https://www.ahri.com.au/media/1162/by-law-1-code-of-ethics-and-professional-conduct_updated-october-2016.pdf

Aloisi, A., & Gramano, E. (2019). Artificial intelligence is watching you at work: Digital surveillance, employee monitoring, and regulatory issues in the EU context. *Comparative Labor Law & Policy Journal*, *41*(1), 95–122.

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal*, *26*(1), 1–11.

Bailey, D., & Barley, S. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization*, *30*(2), 1–12.

Bailie, I., & Butler, M. M. (2018). *An examination of artificial intelligence and its impact on human resources*. CognitionX.

Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., & Zhang, Y. (2018). AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. arXiv preprint arXiv:1810.01943.

Benbya, H., Davenport, T., & Pachidi, S. (2020). Artificial intelligence in organizations: Current state and future opportunities. *MIS Quarterly Executive*, *19*(4), 9–21.

Berg, J. (2019). Protecting workers in the digital age: Technology, outsourcing and the growing precariousness of work. Retrieved January 31, 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3413740

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, *2*(2), 1–9.

Bloodworth, J. (2018). *Hired: Six months undercover in low wage Britain*. Atlantic Books.

Bollen, K. A. (1989). *Structural equations with latent variables* (Vol. 210). John Wiley & Sons.

Bort, J. (2019). Amazon's warehouse-worker tracking system can automatically pick people to fire without a human supervisor's involvement. Business Insider. Retrieved 25 April, 2019, from https://www.businessinsider.com/amazon-system-automatically-fires-warehouse-workers-time-off-task-2019-4?r=US&IR=T

Buckingham, M. (2015). Most HR data is bad data. *Harvard Business Review*. Retrieved December 1, 2021, from https://www.marcusbuckingham.com/wp-content/uploads/2017/08/Most-HR-Data-Is-Bad-Data-HBR.pdf

Byrnjolfsson, E., & Macafee, A. (2014). *The second machine age: Work, progress and prosperity in the time of brilliant technologies*. W.W. Norton.

Callen, A. (2021). When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. *Administrative Science Quarterly*, *66*(4), 1173–1212.

Chamorro-Premuzic, T., Polli, F., & Dattner, B. (2019). Building ethical AI for talent management. Retrieved December 3, 2020, from https://hbr.org/2019/11/building-ethical-ai-for-talent-management

Charlwood, A. (2021). Artificial intelligence and talent management. In S. Wiblen (Ed.), *Digitalised talent management* (pp. 122–136). Routledge.

CIPD. (2020). Code of professional conduct. Retrieved December 18, 2020, from https://www.cipd.co.uk/about/what-we-do/professional-standards/code

Collings, D. G., Nyberg, A. J., Wright, P. M., & McMackin, J. (2021). Leading through paradox in a COVID-19 world: Human resources comes of age. *Human Resource Management Journal*, *31*(4), 819–833.

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020). Biased programmers? Or biased data? A field experiment in operationalizing AI ethics. In *Proceedings of the 21st ACM conference on economics and computation* (pp. 679–681).

CPHR. (2016). Code of ethics and rules of professional conduct. Retrieved December 18, 2020, from https://cphr.ca/wp-content/uploads/2017/01/2016-Code-of-Ethics-CPHR-2.pdf

Crawford, K. (2021). *Atlas of AI*. Yale University Press.

Dattner, B., Chamorro-Premuzic, T., Buchband, R., & Schettler, L. (2019). The legal and ethical implications of using AI in hiring. *Harvard Business Review*. Retrieved December 1, 2021, from https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring

De Stefano, V., & Aloisi, A. (2021). Artificial intelligence and workers rights. Retrieved November 25, 2021, from https://socialeurope.eu/artificial-intelligence-and-workers-rights

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. *Human Resource Management Journal*, *30*(1), 114–132.

Dundon, T., & Rafferty, A. (2018). The (potential) demise of HRM. *Human Resource Management Journal*, *28*(3), 377–391.

Dzieza, J. (2020). How hard will the robots make us work? *The Verge*. Retrieved March 6, 2020, from https://web.archive.org/web/20200319161228/https://www.theverge.com/2020/2/27/21155254/automation-robots-unemployment-jobs-vs-human-google-amazon

EEOC. (1979). Questions and Answers to clarify and provide a common interpretation of the uniform guidelines on employee selection procedures. *Federal Register*, *44*(43).

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280.

Gebru, T. (2021). For truly ethical AI, its research must be independent from big tech. *The Guardian*. Retrieved December 6, 2021 from https://www.theguardian.com/commentisfree/2021/dec/06/google-silicon-valley-ai-timnit-gebru

Glickson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660.

Greasley, K., & Thomas, P. (2020). HR Analytics: The onto-epistemology and politics of metricised HRM. *Human Resource Management Journal*, *30*(4), 494–507.

Green, F., Felstead, A., Gallie, D., & Henseke, G. (2021). Working still harder. *Industrial and Labor Relations Review*. https://doi.org/10.1177/0019793920977850

Grigore, G., Molesworth, M., Miles, C., & Glozer, S. (2021). (Un)resolving digital technology paradoxes through the rhetoric of balance. *Organization*, *28*(1), 186–207.

Guenole, N., & Feinzig, S. (2019). *The business case for AI in HR*. IBM Smarter Workforce Institute.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), e1312.

Hutchinson, B., & Mitchell, M. (2019). 50 years of test (un)fairness: Lessons for machine learning. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 49–58).

Jacobs, A. Z., & Wallach, H. (2021). Measurement and fairness. Retrieved from https://arxiv.org/pdf/1912.05511.pdf

Jatobá, M., Santos, J., Gutierriz, I., Moscon, D., Fernandes, P. F., & Teixeiraade, J. P. (2019). Evolution of artificial intelligence research in human resources. *Procedia Computer Science*, *164*, 137–142.

Jo, E. S., & Gebru, T. (2019). Lessons from archives: Strategies for collecting sociocultural data in machine learning. Retrieved from https://arxiv.org/pdf/1912.10389.pdf

Kellogg, K., Valentine, M., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410.

Kelly, J. (2021). Billion-dollar unicorn and artificial intelligence career-tech startup is improving the job search experience. *Forbes.Com*. 1st December. Retrieved December 3, 2021, from https://www.forbes.com/sites/jackkelly/2021/12/01/billion-dollar-unicorn-and-artificial-intelligence-career-tech-startup-is-improving-the-job-search-experience/?sh=37a869fa4e2b

Kline, R. B. (2015). *Principles and practice of structural equation modelling*. Guilford publications.

Kochan, T. (2021). Artificial intelligence and the future of work: A proactive strategy. *AI Magazine*, *42*(1), 16–24.

Langer, M., & Landers, R. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third party observers. *Computers in Human Behavior*, *123*, 106878. https://doi.org/10.1016/j.chb.2021.106878

Lawrence, T. B., & Suddaby, R. (2006). Institutions and institutional work. In S. Clegg, C. Hardy, T. B. Lawrence, & W. R. Nord (Eds.), *Handbook of organization studies* (pp. 215–254). Sage Publications Ltd.

Lebovitz, S., Levina, N., Lifshitz-Assaf, H., & Lifshitz-Assa, H. (2021). Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly*, *45*(3), 1501–1526.

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute.

Lunden, I. (2021). Beamery raises $138M at an $800M valuation for its 'operating system for recruitment. *TechCrunch*, 17th June. Retrieved December 3, 2021, from https://techcrunch.com/2021/06/17/beamery-raises-138m-for-its-end-to-end-crm-for-recruitment/?guccounter=1

Makarius, E., Mukherjee, D., Fox, J., & Fox, A. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, *120*, 262–273.

Markoff, J. (2016). *Machines of loving grace: The quest for common ground between humans and robots*. Ecco Press.

Matsakis, L. (2019). At an outback steakhouse franchise, surveillance blooms. *Wired.Com*. Retrieved January 18, 2020, from https://www.wired.com/story/outback-steakhouse-presto-vision-surveillance/

Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *Management Information Systems Quarterly*, *45*(4), 1999–2022.

Moore, P. V. (2019). OSH and the future of work: Benefits and risks of artificial intelligence tools in workplaces. In *International conference on human–computer interaction* (pp. 292–315). Springer.

Moore, P. V. (2020). Data subjects, digital surveillance, AI and the future of work. *European Parliament*. Retrieved December 1, 2021, from https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2020)656305

Murray, A., Rhymer, J., & Sirmon, D. (2021). Humans and technology: Forms of conjoined agency in organizations. *Academy of Management Review*, *46*, 552–571.

Narayan, A. (2019). How to recognize AI snake oil. Retrieved January 18, 2020, from https://www.cs.princeton.edu/arvindn/talks/MIT-STS-AI-snakeoil.pdf

O'Connor, S. (2021). Why I was wrong to be optimistic about robots. Financial Times. Retrieved from https://www.ft.com/content/087fce16-3924-4348-8390-235b435c53b2

Office for AI. (2019). Understanding artificial intelligence. Retrieved December 1, 2021, from https://www.gov.uk/government/publications/understanding-artificial-intelligence

O'Mahoney, M., Vecchi, M., & Venturini, F. (2020). Capital heterogeneity and the decline of the labour share. *Economica*, *88*(350), 271–296. https://doi.org/10.1111/ecca.12356

Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*(1), 153–172.

Prikshat, V., Malik, A., & Budhwar, P. (2021). AI-augmented HRM: Antecedents, assimilation and multilevel consequences. *Human Resource Management Review*. https://doi.org/10.1016/j.hrmr.2021.100860

Pyburn, K. M., Jr, Ployhart, R. E., & Kravitz, D. A. (2008). The diversity–validity dilemma: Overview and legal context. *Personnel Psychology*, *61*(1), 143–151.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192–210.

REC. (2021). Data-driven tools in recruitment guidance. Retrieved December 9, 2021, from https://www.rec.uk.com/our-view/research/practical-guides/data-driven-tools-recruitment-guidance#data_protection

Scherer, M. (2017). AI in HR: Civil rights implications of employers' use of artificial intelligence and big data. *SciTech Lawyer*, *13*(2), 12–15.

Scholz, T. M. (2020). Big data and human resource management. In J. S. Pederson & A. Wilkinson (Eds.), *Big data: Promise, applications and pitfalls*. Edward Elgar.

Schulte, B. (2020). Why today's shopping sucks. *Washington Monthly*. January/February/March. Retrieved January 18, 2020, from https://washingtonmonthly.com/magazine/january-february-march-2020/why-todays-shopping-sucks/

SHRM. (2014). Code of ethics. Retrieved December 18, 2020, from https://www.shrm.org/about-shrm/pages/code-of-ethics.aspx

Simonite, T. (2021). What really happened when Google Ousted Timnit Gebru. *Wired*. Retrieved December 1, 2021, from https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/

Singh, M. (2021). AI startup eightfold valued at $2.1B in SoftBank-led $220M funding. *TechCrunch*, 10th June. Retrieved December 3, 2021, from https://techcrunch.com/2021/06/10/ai-startup-eightfold-valued-at-2-1b-in-softbank-led-220m-funding/

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review*, *36*(2), 381–403.

Sonnemaker, T. (2021). Amazon is deploying AI cameras to surveil delivery drivers '100% of the time'. *Business Insider*. Retrieved December 1, 2021, from https://www.businessinsider.com/amazon-plans-ai-cameras-surveil-delivery-drivers-netradyne-2021-2?r=US&IR=T

Strohmeier, S., & Piazza, F. (2013). Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications*, *40*(70), 2410–2420. https://doi.org/10.1016/j.eswa.2012.10.059

Strohmeier, S., & Piazza, F. (2015). Artificial intelligence techniques in human resource management: A conceptual exploration. *Intelligent Techniques in Engineering Management*. *87*, 149–172.

Sundararajan, M., & Najmi, A. (2020). The many Shapley values for model explanation. In *Proceedings of machine learning research 119*. Retrieved December 1, 2021, from https://proceedings.mlr.press/v119/sundararajan20b.html

Susskind, D., & Susskind, R. (2015). *The future of the professions: How technology will transform the work of human experts*. Oxford University Press.

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resource management challenges and a path forward. *California Management Review*, *61*(4), 15–42.

Teodorescu, M. H. M., Morse, L., Awwad, Y., Kane, G. C., & Kane, G. (2021). Failures of fairness in automation require a deeper understanding of human-Ml augmentation. *Management Information Systems Quarterly*, *45*(3), 1483–1500.

Thomas, R. J. (1994). *What machines can't do: Politics and technology in the industrial enterprise*. University of California Press.

Thompson, P. (2010). The capitalist labour process: Concepts and connections. *Capital & Class*, *34*(1), 7–14.

Ton, Z. (2012). Why good jobs are good for retailers. *Harvard Business Review*, *90*(1–2), 124–131.
van den Broek, E., Sergeeva, A., Huysman, M., & Huysman Vrije, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *Management Information Systems Quarterly*, *45*(3), 1557–1580.
Veale, M., & Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act. *Computer Law Review International*, *22*(4), 97–112.
Von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, *4*(44), 404–409.
Whittaker, M. (2021). The steep cost of capture. *ACM Interactions*, *XXVIII*(6), 50–55.
Williams, M., Zhou, Y., & Zou, M. (2020). *Mapping Good Work: The quality of working life across the occupational structure*. Bristol University Press.
Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. *New Labor Forum*, *28*(1), 10–29.

How to cite this article: Charlwood, A., & Guenole, N. (2022). Can HR adapt to the paradoxes of artificial intelligence? *Human Resource Management Journal*, 1–14. https://doi.org/10.1111/1748-8583.12433