ORIGINAL PAPER

AI recruitment algorithms and the dehumanization problem

Megan Fritts[1] 🅖 · Frank Cabrera[2]

Abstract

According to a recent survey by the HR Research Institute, as the presence of artificial intelligence (AI) becomes increasingly common in the workplace, HR professionals are worried that the use of recruitment algorithms will lead to a "dehumanization" of the hiring process. Our main goals in this paper are threefold: (i) to bring attention to this neglected issue, (ii) to clarify what exactly this concern about dehumanization might amount to, and (iii) to sketch an argument for why dehumanizing the hiring process is ethically suspect. After distinguishing the use of the term "dehumanization" in this context (i.e. removing the human presence) from its more common meaning in the interdisciplinary field of dehumanization studies (i.e. conceiving of other humans as subhuman), we argue that the use of hiring algorithms may negatively impact the employee-employer relationship. We argue that there are good independent reasons to accept a substantive employee-employer relationship, as well as an applicant-employer relationship, both of which are consistent with a stakeholder theory of corporate obligations. We further argue that dehumanizing the hiring process may negatively impact these relationships because of the difference between the values of human recruiters and the values embedded in recruitment algorithms. Drawing on Nguyen's (in: Lackey, Applied Epistemology, Oxford University Press, 2021) critique of how Twitter "gamifies communication", we argue that replacing human recruiters with algorithms imports *artificial values* into the hiring process. We close by briefly considering some ways to potentially mitigate the problems posed by recruitment algorithms, along with the possibility that some difficult trade-offs will need to be made.

Keywords Hiring Algorithms · Employee-Employer Relation · Stakeholder Theory · Gamification · Dehumanization · Human Resources

Introduction

As the presence of artificial intelligence (AI) becomes increasingly common in the workplace, Human Resources (HR) professionals are worried that the use of recruitment algorithms will lead to a "dehumanization" of the hiring process. In this paper, our aim is to examine this concern, which has received little attention in debates about the ethics of algorithms. We begin by discussing a recent survey of HR professionals, which reports current attitudes about the use of such AI-based tools in the workplace. In general, expectations are such that in the next few years, AI will play a prominent role in the HR toolkit, especially for hiring and onboarding purposes. Perhaps the most common objection to the use of hiring algorithms or algorithmic decision-making systems in general is that they have the potential to be biased or lead to objectionable patterns of discrimination. However, while HR professionals have registered concerns about bias and discrimination, interestingly the *most cited* worry in the recent survey that we consider is that hiring algorithms will "dehumanize" the hiring process. Our main goals in this paper are threefold: (i) to bring attention to this neglected issue, (ii) to clarify what exactly this concern about dehumanization might amount to, and (iii) to sketch an argument for why dehumanizing the hiring process is ethically suspect. After distinguishing the use of the term "dehumanization" in this context (i.e. removing the human

Megan Fritts and Frank Cabrera have contributed equally.

✉ Megan Fritts
   mfritts1@css.edu

   Frank Cabrera
   cabrera@msoe.edu

[1] Philosophy Department, The College of St. Scholastica, Duluth, MN, USA

[2] Humanities, Social Science, and Communication Department, Milwaukee School of Engineering, Milwaukee, WI, USA
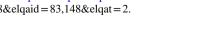
presence) from its more common meaning in the interdisciplinary field of dehumanization studies (i.e. conceiving of other humans as subhuman), we argue that fears about dehumanizing the hiring process can be fruitfully investigated by considering its potentially negative impact on the employee-employer relationship. We argue that there are good independent reasons to accept a genuine, substantive employee-employer relationship, as well as an applicant-employer relationship, both of which are consistent with a stakeholder theory of corporate obligations. We go on to argue that dehumanizing the hiring process may negatively impact these relationships because of the difference between the values of human recruiters and the values embedded in recruitment algorithms. Drawing on Nguyen's (2021) critique of how Twitter "gamifies communication", we argue that replacing human recruiters with algorithms imports *artificial values* into the hiring process. We close by briefly considering some ways to potentially mitigate the problems posed by recruitment algorithms. However, as the use of recruitment algorithms becomes more widespread, difficult trade-offs may need to be made, since the human element, as it features in the hiring process, can also be the source of its own problems.

HR managers react to the idea of hiring algorithms

In 2019 the HR Research Institute[1] published a report entitled "The 2019 State of Artificial Intelligence in Talent Acquisition."[2] This report was based on a survey of 484 responses from HR professionals across a wide variety of industries and includes representatives from both small businesses and large firms. According to the report "many HR professionals foresee AI playing enhanced roles in most aspects of talent acquisition" over the next few years (3). There are many ways in which AI may contribute to HR's traditional functions, e.g. chatbots to answer employee questions and to guide recently hired employees through the onboarding process. But according to *recruitment* professionals, what they desire most from the use of AI is to "find more qualified candidates" and "to reduce the time they spend sifting through resumes" (4). One reason for interest in utilizing AI for recruitment is that only about 25% of survey respondents rated their talent acquisition methods very favorably (6). Even though currently only 10% of survey

respondents claimed that their company uses AI to a high or very high degree in the hiring process, according to the report, 36% of respondents claimed that in the next two years they expect their company to begin making heavy use of AI for recruitment, and 61% expect to see at least moderate usage (11).

While survey respondents expect that *every* stage of the hiring process will be impacted by AI in the next few years, some stages are expected to be more affected than others. The first stage of the recruitment process consists of *sourcing* a pool of potential candidates. For lucrative positions, a firm might not need to do much advertising to attract applicants. Still, we can imagine that AI-based tools might be used to conduct *targeted searches* to come up with a preliminary pool of potential candidates (14–18). The second stage of the hiring process involves *screening* potential candidate to narrow down what might be a rather large pool. Here, AI-based tools might be used to quickly process and prioritize resumes or to predict the fit between a candidate's personality and the organization's "culture" (19–29). The third stage of the hiring process is the *interview stage*, where candidates are evaluated more closely by the hiring committee. Here, AI-based tools might be used to automate the first round of interviews or to contribute to the evaluation of candidates (30–4). The last stage of the hiring process is the *selection stage*, where an offer is extended to the successful applicant. Here, AI-based tools might influence the final selection process by recommending the best candidates in the remaining pool (34–7). Of course, even if computational algorithms directly impact the selection stage, it will prove some time before AI comes close to having the final say in hiring decisions. In general, survey respondents expect AI to play the largest role in the earlier phases of the recruitment process, e.g. sorting through resumes to establish a preliminary pool, since these are the facets of the hiring process that depend less on sophisticated human judgment calls, and so are more amenable to automation (12).

To get a better sense of how AI is currently impacting the hiring process, consider the empirical study by Raghavan et al. (2019) of 18 vendors offering AI-based recruitment services that are currently being employed by HR professionals at the *screening* stage. Raghavan et al. identify various assessment types that are used by these vendors, including personality tests, situational judgment tests, gameplay (e.g. puzzle-solving), and video interview analysis (2019, p. 6). These assessment tools can be either customized to meet the needs of the client, or they can be pre-built and general-purpose. In conjunction with "training data" provided either by the vendor or the employer, proprietary algorithms are used to make predictions about the expected degree of success of a prospective employee. For example, HR might provide one of these AI recruitment vendors with data on past employees and request a

---

[1] The HR Institute is the research division of HR.com, the largest online community for HR professionals.

[2] https://www.oracle.com/a/ocom/docs/artificial-intelligence-in-talent-acquisition.pdf?elqTrackId=1279a8827f3d4548ae3f966beeeef458&elqaid=83,148&elqat=2.

report that predicts the expected sales numbers or performance reviews of the prospective employee based on certain features about them. To do so, the vendor might use computational algorithms to analyze a short video of the candidate that detects desirable personality traits thought to be correlated with success, such as "openness", "enthusiasm", and "warmness". As Raghavan et al. note, these "video-based assessments, in particular, are increasingly common" (2019, p. 12). Indeed, the practice has become widespread enough to draw the attention of lawmakers. In 2019 Illinois, USA passed the *Artificial Intelligence Video Act*, which requires employers to inform and obtain consent from applicants if their video interviews will be analyzed by some AI-based assessment tool.

The dehumanization problem

The most cited ethical concern in the HR research institute report

Many scholars across a wide array of disciplines have raised a host of ethical concerns about the use of predictive analytics algorithms for making decisions that greatly affect the life prospects of human beings (Mittelstadt et al., 2016). Perhaps the most common objection that the use of such technology prompts is that it already is or has the potential to be *biased* or *unfair,* leading to objectionable patterns of discrimination (e.g. Eubanks, 2018; Noble, 2018). A particularly high-profile example of a biased hiring algorithm was one built by Amazon (but allegedly never used), which systematically discriminated against resumes from women applicants (Dastin, 2018). While HR managers have registered worries about bias and discrimination, interestingly, *the most cited* ethical concern about the use of hiring algorithms in the HR Research Institute report is the "dehumanization of the recruitment process" (49). Unfortunately, the report does not elaborate much on how we should understand "dehumanization" in this context. Remarks along similar lines include the concern that the use of AI for hiring purposes will make the recruitment process lose its "human touch" (49). This ethical concern about the possibility of dehumanizing the hiring process is interesting because it does not depend on the outcome of debates over whether hiring algorithms are biased. Future technical and ethical advances may address problems of bias and discrimination (Kleinberg et al., 2020), but worries about dehumanization seem to be largely independent of these other issues. Since the worry about dehumanizing the hiring process was the ethical concern that was most widely cited in this survey of HR professionals, it is worthwhile to attempt to get clearer on the nature of the problem.

Distinguishing two distinct meanings of "dehumanization"

As Smith (2018, p. 264) aptly observes, "mentioning dehumanization raises the question of what dehumanization is. That question is surprisingly difficult to answer." In trying to make sense of the worry that hiring algorithms will dehumanize the recruitment process, one might attempt to investigate the phenomenon in light of the interdisciplinary field of dehumanization studies, which spans several disciplines including philosophy, political science, social psychology, disability studies, history, etc. The modern origin of dehumanization studies is often traced back to scholarly reactions to the atrocities of the twentieth century, especially Hannah Arendt's work on the social mechanisms by which the Holocaust was perpetrated (Kronfeldner, 2021, p. 3). Given that the field of dehumanization studies is "rather patchy" (Kronfeldner, 2021, p. 1), it comes as little surprise that there are many different accounts of dehumanization in the literature. As Smith (2016, pp. 418–19) helpfully points out, there are at least eight different ways in which "dehumanization" has been conceptualized by various scholars:

  (i) Subjecting others to indignities; or, in a more Kantian vein, treating them merely as means.
 (ii) Verbally likening others to nonhuman animals or inanimate objects.
(iii) Denying the subjectivity, individuality, agency, or distinctively human attributes of others.
 (iv) Denying that others undergo mental states.
  (v) Treating others in such a way as to erode, obstruct, or extinguish some of their distinctively human attributes.
 (vi) Conceiving of others as inanimate objects.
(vii) Conceiving of others as less human than members of one's ingroup.
(viii) Conceiving of others as subhuman creatures.

The sense of "dehumanization" that Smith favors and elaborates upon (e.g. 2014, 2016, 2018) is that which is suggested by (viii), conceiving of others as subhuman, or somewhat paradoxically, as both human and subhuman simultaneously. Many of the other senses of dehumanization listed above are related to and often go hand-in-hand with the process of conceiving of others as subhuman, but the latter, as Smith argues, is what is essential to dehumanization (2016, p. 419).

Although there are many different ways in which dehumanization might manifest itself, the idea that dehumanization, at bottom, involves the conception of other humans as less than human and thus lacking moral standing, captures the core element of dehumanization. As Maria Kronfeldner observes in her valuable introduction to the *Routledge*

Handbook of Dehumanization, despite its many complexities, in cases of dehumanization "one aspect seems to be always present: dehumanization establishes *difference* and *distance* between human beings (2021, p. 9). For example, the atrocities inflicted on the Jewish people and other targeted groups by the Nazis during World War II "are often taken as paradigmatic cases of dehumanization", and of course the Nazis explicitly conceived of the victims of their brutality "as less human if not less than human" (Kronfeldner, 2021, p. 14). The ability to dehumanize other human beings, in the sense of conceiving of them as subhuman in this way, is made possible, as Smith argues, by "our propensity for essentialistic and hierarchical thinking" (2018, p. 268).

Unfortunately, as our brief sketch of standard accounts of dehumanization makes clear, the scholarly literature on dehumanization will not be of much help in clarifying the central concern that HR managers have about hiring algorithms. Despite the use of term "dehumanization" in the HR Research Institute report, it is implausible that survey respondents have in mind the standard sense of dehumanization whose essence involves conceiving of other human beings as subhuman. Clearly, nothing so extreme as the phenomenon that dehumanization scholars investigate is liable to occur in the case of hiring algorithms. Moreover, given that this form of dehumanization is often associated with grave human rights violations and even genocide, it would be offensive to even make that comparison. But if we cannot understand the qualms that HR managers have about hiring algorithms by reference to the most prominent accounts of dehumanization, then one might be suspicious of the concerns raised by the HR Research Institute report. More specifically, one might worry that the term "dehumanizing" is being employed here, as is often the case in conversations about new technologies, as "an unspecific negative qualifier—used in order to stress a negative evaluation of whatever phenomenon is under study" (Kronfeldner, 2021, p. 17). This worry is particularly pressing given that questions about whether new technologies are dehumanizing have been denounced by some critics as "largely meaningless" because, it is claimed, terms like "dehumanizing", when not referring to the phenomenon investigated in dehumanization studies, are "not well defined" (Mowshowitz, 2008, p. 281).

While we think that the worries that HR managers have cannot be subsumed under the most prominent accounts of the phenomenon investigated in the dehumanization studies literature, we do not think these concerns should be dismissed so easily. Certainly, the claim that hiring algorithms have the potential to be dehumanizing is not literally *meaningless*. The sense of dehumanization at play in the case of hiring algorithms does not involve conceiving of other humans as subhuman, but instead involves removing the human presence from some process or domain, where previously the human presence was taken as a given. This sense of dehumanization is not unique to the burgeoning debate over hiring algorithms. We can find similar uses of the term "dehumanization" in the debate over "lethal autonomous weapons systems", e.g. drone warfare. For instance, after acknowledging that the term "dehumanization" can refer to the "war victims' sufferings", Joerden (2018, p. 56) notes "there is a meaning of the term 'dehumanization' that must be set apart, namely, that mankind withdraws, so-to-speak, from the immediate war-related events, and thus, 'dehumanization' of war occurs." We might also refer to this alternate meaning of dehumanization succinctly as "leaving out the human" (Kronfeldner, 2021, p. 17). Henceforth, this is the account of dehumanization we will have in mind. With this alternative sense of the term "dehumanization" in place, it is clear that hiring algorithms would leave out the human, and thus it is clear that hiring algorithms would dehumanize the hiring process.

The central question now though is whether there is anything wrong with leaving out the human in this particular context. It requires little argument to show that conceiving of and treating other humans as subhuman is always seriously morally wrong. Indeed, this more common sense of dehumanization investigated by Smith and others seems to be a negatively valenced instance of what Bernard Williams (1985) would call a "thick ethical concept." These are concepts that already have evaluative content "built into" them. The common sense of "dehumanization" clearly already has wrongness built into it. However, as we have argued, this is not the sense of dehumanization at issue in the case of hiring algorithms, which instead involves removing the human presence. Moreover, this alternative sense of dehumanization is *not* a thick ethical concept; it does not have wrongness built into it. In some cases, it is morally permissible to leave out the human, e.g. having a robot vacuum the floor rather than a human. The problem then though is that once we distinguish these two different senses of the term "dehumanization," it is not at all obvious that dehumanizing the hiring process, in the sense of removing the human presence, is morally objectionable. In the following sections, we aim to give a novel argument, drawing on some related discussions in the ethics of technology, for why dehumanizing the hiring process in this alternative sense *should* be regarded as ethically suspect.[3]

---

[3] We think our formulation of this alternative sense dehumanization in terms of leaving out the human is sufficiently clear for our purposes, but we do not pretend that this is a completely unproblematic conceptual analysis. For example, some questions about how much the human presence must recede in order to count as dehumanizing the process will remain. Even in the debate on lethal autonomous weapons systems, humans have withdrawn a considerable degree from the process, but of course, not entirely so. Presumably, how much withdrawal is sufficient to count as a dehumanized process will depend on the context.

A threat to the employee-employer relationship

The shareholder theory versus the stakeholder theory

A fruitful approach for better understanding why dehumanizing the hiring process might be ethically suspect is to consider how the use of recruitment algorithms might negatively impact the *employee-employer relationship*. Now, those in the grips of Friedman's (1970) infamous "shareholder theory", which states that the sole obligation that corporate executives have is to run the firm in accordance with the wishes of the shareholders, which "generally will be to make as much money as possible while conforming to the basic rules of the society", might balk at the idea of a genuine, substantive employee-employer relationship. The shareholder theory, however, has been the object of much criticism, which continues into the twenty-first century (e.g. Aglietta & Rebérioux, 2005; Stout, 2012). As a result, an alternative to Friedman's model of corporate governance has gained prominence in subsequent decades—a view often referred to as the "stakeholder theory" (e.g. Freeman, 1984).[4] Contra Friedman, the stakeholder theory states that the corporation has obligations to *all stakeholders*, whether or not this maximizes profit. In the widest sense, stakeholders include any groups or individuals "who benefit from or are harmed by, and whose rights are violated or respected by, corporate actions" (Freeman, 1984, p. 100). Typically, this includes owners, employees, suppliers, contractors, customers, investors, local communities, the media, the government, etc.

Nowadays, it is becoming increasingly common for corporate executives to at least pay lip service to the stakeholder theory. In August 2019, the "Business Roundtable" released a statement signed by almost 200 hundred corporate executives, including the CEOs of Amazon, Apple, JPMorgan Chase, Disney, etc., explicitly disavowing previous shareholder-centric principles of corporate governance, and instead pledging "commitment to all of [their] stakeholders.".[5] With respect to employees specifically, the signers pledge to: (i) compensate employees fairly and provide them with important benefits, (ii) support them with training and education to develop new skills, and (iii) promote diversity, inclusion, dignity, and respect in the workplace. Since the stakeholder theory does not subordinate all corporate actions and values to profit maximization, the theory allows for a richer and more robust conception of the ideal employer-employee relationship.

Individuating and distinguishing distinct social relationships

Beyond its connection to the stakeholder theory, there are good independent reasons to accept a substantive conception of the employee-employer relationship. In an influential discussion of the value of privacy, James Rachels argues that privacy is important for "our ability to create and maintain different sorts of social relationships with different people" (1975, p. 326). In doing so, Rachels offers us some helpful clues for how to establish the reality of distinct social relationships. As Rachels observes, distinct social relationships are associated with different patterns of behavior; for example, "a man may be playful and affectionate with his children…but businesslike with his employees, and respectful and polite with his mother-in-law" (ibid.). Moreover, the information that it is appropriate for a person to share depends on the nature of the social relationship at stake. Crucially, these distinct patterns of behavior, along with what information it is appropriate for others to have, are also that which *individuate* and *distinguish* distinct social relationships. As Rachels (1975, p. 327) points out, "different patterns of behavior are (partly) what define the different relationships; they are an important part of what makes the different relationships what they are." Although there may be some cultural and individual variation in one's conception of these distinct relationships, "there is inseparable from that conception an idea of how it is appropriate to behave with and around them, and what information about oneself it is appropriate for them to have" (Rachels, 1975, p. 329). Indeed, more specific discussions of privacy and the employee-employer relationship presuppose something akin to Rachels' general thesis that norms about the sharing of information are inextricably connected to the distinct nature of the social relationship at issue (Simms, 1994; Superson, 1983).

The upshot for our purposes is that we know that there is a genuine, substantive employee-employer relationship because we know that there exist here different patterns of behavior and different standards of privacy, when compared to, say, a friendship or the generic benefactor-to-beneficiary relationship. In light of the dominance of the stakeholder theory of corporate obligations and the fact that there are good independent reasons to believe in a genuine, substantive employee-employer relationship with its own distinct norms, we have good reason to care about the potential impact of hiring algorithms on the relationship between employees and their employers.

In addition to the employee-employer relationship, the foregoing considerations should also make us concerned

---

[4] For a more recent discussion of the stakeholder theory, see Freeman et al. (2018).

[5] See: https://opportunity.businessroundtable.org/ourcommitment/.

about the relationship between employers and applicants. As is the case with other distinct social relationships, unique patterns of behavior and what information it is appropriate to share are associated with the applicant-employer relationship. Since, typically, at most a few candidates are hired from a pool of sometimes hundreds of applicants, it is noteworthy that the majority of those who will be affected, and potentially wronged, by hiring algorithms in any *given* search will be applicants. Thus, if we take a stakeholder perspective, we should care about the impact of hiring algorithms not just on the employee-employer relationship, but also on the relationship between applicants and employers. For convenience, we will speak of the "employee-employer" relationship, but our argument will apply, *mutatis mutandis*, to the prospective-employee-employer relationship.

Artificial values in recruitment algorithms

In the section "The dehumanization problem", we drew attention to the primary concern that HR managers have with recruitment algorithms: the worry that such a hiring method would *dehumanize* the recruitment process. In the section "A threat to the employee–employer relationship", we proposed to understand ethical qualms about dehumanizing the recruitment process (i.e. removing the human presence) as fears about negatively impacting the employee-employer relationship, although it is still unclear exactly what sort of violation recruitment algorithms would constitute. In this section, we aim to further elaborate on why it is that dehumanizing the recruitment process by the use of hiring algorithms poses a threat to the employee-employer relationship.

Gamification and artificial value creation

To understand how the use of hiring algorithms has the potential to negatively impact the employee-employer relationship, we can draw inspiration from criticisms of other, similar moral conundrums related to new and emerging technologies. In his paper "How Twitter Gamifies Communication" C. Thi Nguyen (2021) argues that Twitter "gamifies" communication between people, changing the nature of their discourse. By "gamifies communication", Nguyen (2021, p. 411) specifies that Twitter offers, "immediate, vivid, and quantified evaluations of one's conversational success" in a way that allows the user to improve upon past "performances" (tweets) to score more "points" (likes, retweets, and followers). Of course, gamification is not always bad—in fact, it is often a good thing, allowing us to complete tasks more easily and more enjoyably. As Nguyen points out, (2021, pp. 413–414), we can use game-based learning

to teach ourselves healthy habits such as exercise and proper nutrition. For example, devices like *FitBit* allow users to track their health progress in ways that reward them for meeting fitness goals; apps such as *Calm* even gamify relaxation, by rewarding the user for participating in exercises that are meant to induce physical and mental relaxation.

The problem with the gamifying *sincere* interpersonal communication though ultimately stems from how games motivate participation. Crucially, Nguyen argues, games draw us in by *simplifying* and thus *clarifying* our desires: "Gamification increases our motivation by changing the nature of the activity. Often, the goals of ordinary activity are rich and subtle. When we gamify these activities, we change those goals to make them artificially clear" (2021, p. 411). For example, when engaged in non-gamified *sincere* discourse with our peers, we have complex and multifaceted values for this complex and multifaceted good, as (presumably) ought to be the case. But to "succeed" at Twitter, one must engage in discourse with an *artificial* value set—the set given to them by the site creators—and focus on the goal of the game: maximizing engagement with their tweets. There may be many tasks which could be unproblematically gamified, in virtue of the fact that we don't believe we ought to hold rich, subtle, complex values for them. For instance, toothbrushes which gamify oral hygiene for children may fall into this category.

For Nguyen, values are "artificial" when they are a part of a structure of agency that has been given to us by someone else. In games, someone is the game designer who "sculpts the temporary agency that the player will occupy during the game", determining "who the player will be" in the imagined world of the game (Nguyen, 2021, p. 414). Artificial values, then, are those values that are part and parcel of the entire agential structure that we are given—our end goals, our character, the ways in which we are to move and act, etc.—either by the designer of the game (if we are playing a game) or by some other mechanism that suspends our usual agency.

Furthermore, such suspensions of autonomous agency grant the same artificial values to all persons whose agency is suspended by the same person or mechanism. This is what allows for multi-player games and what allows Twitter to function as efficiently as it does. But for people to share values, Nguyen argues, these values must be simpler and clearer than our ordinary values are outside the context of a pre-sculpted agency. In real life, our values are often opaque to us, and figuring out how to balance them against one another (e.g. "I want to relax tonight but I also want to support my friend by going to her art show") is anything but clear. But in a game, our values are perfectly clear, which allows us to effectively strategize. Nguyen uses the example of our values for "likes", "retweets", and "followers" handed to us by the site designers when we enter the "game" of

Twitter. Of course, outside of Twitter, where our agency is not "sculpted" for us, we still value being liked and admired by our peers. But valuing peer admiration, popularity, or even fame, is not the same thing as valuing Twitter likes or retweets, which are specific discrete digital events. On Twitter, interaction with one's tweets is the entire point of tweeting, and tweets are designed and hashtagged with maximum engagement as the goal. Outside of "game play", our value for peer recognition is simply one value among a host of innumerable social values, and maximum attention is *not* the goal we generally have when interacting with others in the real world outside of Twitter.

Similarly, in board games like tic-tac-toe, we adopt an artificial value for getting three X's or O's in a row. These are artificial values because i) they are values we do not hold outside of the game, and ii) they are simple and clear—we are never conflicted about whether we should try to attain three X's or O's in a row. In general, when we engage in a game or a gamified activity that is well-structured, we are never in the dark about how important certain goals or achievements are. Our artificial values are crystal clear to us because the game designer hands them to us, perfectly packaged. This is what makes game-play fun instead of laborious. "Gamifying" an otherwise difficult or boring task, therefore, is an attempt to "impose value clarity on a pre-existing thicket of values" (Nguyen, 2021, p. 415).

Of course, there is a clear difference between the gamification of Twitter and the ways in which recruitment algorithms are currently being employed by some firms. For one thing, Twitter intentionally gamifies communication by handing the user a set of artificial values that will keep them hooked on the site. Recruitment algorithms usually do not intentionally gamify the hiring process, although there are notable exceptions, as in the case of game-based assessments discussed above. Our primary argument in this section is *not* that recruitment algorithms may dehumanize the hiring process by gamifying it. We introduce the concept of gamification only because this is how Nguyen introduces his concept of artificial values. But artificial values may enter into an activity even if that activity has not been fully gamified, and this, we will argue, makes the use of recruitment algorithms *prima facie* morally objectionable.

Artificial values and recruitment algorithms

Although gamification of the hiring process per se does not entirely justify concerns about dehumanization, we believe both gamification and the use of AI recruitment algorithms share a common worrisome feature. Our proposal here is that the same problem that Nguyen sees with gamified communication arises in the case of offloading hiring responsibilities to recruitment algorithms, namely, the creation of *artificial values* in human-to-human relationships. Recall

Nguyen's critique of Twitter's gamification of communication is that it involves a sort of trade: our normal, complex, opaque, human values, in exchange for easy artificial ones. When we play games, or are involved in gamified activities, our practical agency is ruled by these artificial values, rather than our normal values that we act on in our real lives. The concern in the case of gamified communication is that the complexity of how and why we value human communication *ought not* to be simplified.

In the employment context, how these artificial values arise is rather simple: algorithms are quantifiers, and as such will make hiring decisions or recommendations through a process of quantification based on the values fed into the system by the programmer. As we discussed in the section "HR managers react to the idea of hiring algorithms", the vendors that sell these assessment tools are quick to assure us that this quantified predictive process is incredibly complex and extremely accurate, based on careful validation studies, etc. Still, as AI currently exists, even the most complex algorithmic values are not identical to human values—they are artificial, created by the human programmer. Something that the use of AI-based tools for recruitment purposes highlights is that human relationships look nothing like that which holds between a person and a computer algorithm. This is, in part, because the values of the recruitment algorithm are never affected by the actual human behind the resume. In the context of human relationships, we expect to be appreciated in the sort of way humans appreciate other humans, rather than be understood through a quantified assessment of individual bits of personal data.

To better understand the concern with AI algorithm values replacing the real human values in hiring situations, it will help us to imagine a very value-sensitive recruitment scenario. Imagine a ballet dancer who has been training since childhood to audition for the New York City Ballet. As she sits in the overflow room, waiting for her turn to audition, she is informed that the judges will only see two more dancers before they have to go home for the evening, and that she will not be one of those two. She is told, however, not to worry. The judges have purchased a highly advanced algorithmic decision-making system that collects millions of points of data on all remaining contenders, scans film of their dances in search of skills that predict future success, and even checks their family history for relatives who have struggled with substance abuse or obesity. The dancer is heartbroken. But why?

There are two reasons the dancer may be dismayed to learn that her dancing will be judged by an algorithm rather than the human judges she had anticipated. The first reason is that she may feel she has less chance of success being judged by the algorithm than did the dancers who performed for the human judges. We anticipate many readers will connect with this feeling, though it is hard to put a finger on

exactly why this is. Why should our ballet dancer *not* be relieved to be performing for such an *impartial* judge, capable of assessing her raw talent without being distracted by biases or illicit expectations? The answer seems to be that we are often more confident in our ability to communicate important aspects of ourselves to humans rather than to computers. Most of us have been inculcated with skills of human interaction which run deeper than any purely computational or theoretical analysis could explain; we are able to read people, inspire people, reassure and calm people, and communicate with people through untold numbers of subconscious movements, facial expressions, verbal and non-verbal vocal cues, along with too many others to list here. With AI we are, to put it simply, out of our element. It is not how we naturally communicate, and this puts us at a disadvantage.

The second reason is more along the lines of our primary critique of recruitment algorithms: the dancer may worry that, even if she gets hired, the victory will be hollow. Rather than being chosen because her dancing truly and deeply moved the judges, she will have gotten the job because thousands of details about her were quantified and compared to the traits of others, and hers came out on top. While she may be comparably happy to be hired, she is unhappy about how and why she is hired, because she feels reduced to her attributes, disjointed, and objectified. Or, perhaps more accurately, she feels cut up into small bits of data, unsure whether herself as whole dancer—a whole person—is truly wanted as an employee of the ballet. Already we can see here a fraught element introduced into the would-be employer-employee relationship.

There is a more precise way of describing what seems morally objectionable about auditioning for an algorithm. Like Nguyen's critique of Twitter's gamification of communication, we suggest that in the case of the ballet dancer, the values placed on the skills of the dancers are given an *artificial clarity* by an algorithm that is programmed to do the job of the human judges. No matter how hard the programmer tries, the values of the algorithm can never really be like the values of those particular judges. No matter how hard *she* tries, the ballet dancer's dancing will not—cannot—be valued by the AI in the same way that it would be valued by human judges. It is not, to be clear, merely in virtue of the program's status as AI that it cannot value the dancer in the same way that a human judge would. The AI's values are artificial because i) they are given to the AI by another entity (its programmer) and ii) because they enjoy a clarity that distinguishes their metric of weighting and comparing from the opaque metric of regular human values. Nguyen elaborates on this concern as follows:

> In ordinary life, our values are hard to balance. I care about spending time with my loved ones, raising my children right, writing good philosophy, enjoying

myself in rock climbing, staying healthy, and eating delicious food. Not only are my values often in tension, but there is usually no way to precisely compare them. (2021, p. 415)

This seems right—any attempt to articulate precisely how our values affect each other, weigh and pull against one another to guide our everyday lives, will be, strictly speaking, a falsehood. If the judges value things like grace, strength, facial expressiveness, fluidity, genuineness, and earnestness, it is anyone's guess how exactly each of these values will be weighted in the individual performers. For any two dancers with equal amounts of grace and earnestness, the earnestness of one may put her ahead of the other for no other reason than that the judges value her *personal earnestness* in ways too complex and opaque to ever elaborate on. Likewise, AI recruitment algorithms rely entirely on artificial value trade-offs that rob our human relationships of what makes them distinctively valuable—perhaps, that our human relationships are not very much like a precisely-programmed algorithm at all. There are obviously disanalogies between the job of a ballet dancer and the sort of job AI recruitment algorithms would normally be recruiting for—but as we see it, the difference is one of degree, not kind.[6]

Is the employee–employer relationship worth saving?

The sorry state of the employee–employer relationship

Thus far we have been motivated to better understand the most widely cited concern among HR managers regarding the increasing use of predictive algorithms in the hiring process. More than anything else, HR managers are worried that the process of identifying candidates, screening and interviewing them, and onboarding them, will lose the "human touch" if conducted by AI. As we have proposed, one reason to view this sort of dehumanization (i.e. removing the human presence) as ethically suspect is that the artificial values imported by hiring algorithms may be a threat to the employee-employer relationship, in the sense that the relationship is stripped of something characteristic of the ways in which humans typically engage with one another:

---

[6] Our argument in this section should be viewed as part of an ongoing conversation about some of the detriments of advanced technology. This conversation was spearheaded in the twentieth century by some classical critiques of technology, industrialized society, and "instrumental rationality", often associated with authors belonging to the Frankfurt School, such as Marcuse (1941), Adorno and Horkheimer (1972/2020), and Habermas (1970). See Schecter (2010) for a general overview.

the complex, opaque ways in which the values appear and interact in the relationship. But is the employee-employer relationship worth saving? One may object that, even if hiring algorithms posed a threat to the employee-employer relationship in this way, the relationship is already fraught with so many non-ideal aspects that improving—or preventing further damage to—the relationship would be wasted effort.

To be sure, business ethicists and political philosophers alike have long bemoaned the sorry state of the contemporary employee-employer relationship, even before the rise of AI in the workplace. A point that Borowski observed over two decades ago would ring true to many employees today, namely that the relationship between the employee and the employer "seems to be an adversarial one—almost as if management and workers were on two different competing teams" (1998, p. 1624). Indeed, this is precisely how Joseph Jones, a worker in an Alabama Amazon warehouse recently described his experience: "It's a very adversarial relationship with the supervisors and the staff…You're a cog in the system…and it's very obvious" (Sherman, 2021). An adversarial relationship of this sort is liable to lead to feelings of tension, distrust, and antagonism (Borowski, 1998, pp. 1625–6).[7] In a similar fashion, Karnes (2009, p. 189) laments the "disintegration of employer-employee relationships." According to Karnes (2009), gone are the days when workers and employers trusted each other (190), when both parties demonstrated commitment to each other's well-being (191), when employees and employers genuinely felt a sense of unity and togetherness (192), etc. These impressions are supported by some empirical data. According to the 2014 "Work and Well-Being Survey" conducted by the *American Psychological Association*, "employee distrust is pervasive in the U.S. workforce."[8] More specifically, only 50% of workers believed that their employer is "open and upfront" with them, and about 25% of workers claimed that they do not trust their boss.

Some modest proposals

To counteract the growing sense of distrust, tension, and antagonism between employees and employers, Borowski proposes injecting a dose of Kantian ethics into the workplace: both workers and employers need to treat each other with the respect that any human person deserves in virtue of

being a rational and autonomous creature (1998, p. 1626–7). A similar solution put forward by Karnes consists, largely, in putting the stakeholder theory into practice: employers should take into account workers' basic human needs, seek to create enthusiastic employees, treat their workers fairly and justly, recognize the genuine accomplishments of their employees, and foster genial and cooperative relations in the workplace (2009, pp. 194–5). In general, employers and employees ought to treat each other with the decency and respect that each party is owed in virtue of being human, regardless of profit margins (Karnes 2009, p. 195). We might take some inspiration from these proposals in addressing the issue of using hiring algorithms in the workplace. Insofar as employers opt to use algorithms for hiring purposes, then they should, as far as possible, be guided by these broadly Kantian ideals of respect for persons *as such*, which may involve limiting the use of hiring algorithms to, say, the *screening stage*. This is not to say that employers must themselves be Kantians. Such a principle is available to those holding a wide array of normative views, including ethical pluralism (Ross, 1930), utilitarianism (Mill 1864/2014), virtue ethics and care ethics (Kittay, 2005). Most of these frameworks take respect for persons as a *prima facie* value, though different frameworks will have different reasons for doing so.

In our view then, even though the state of the employee-employer relationship may already be quite bad, this does not mean that we ought not to care about the impact of the use of hiring algorithms on the relationship between employees and their employers. To the extent that HR managers are themselves concerned about dehumanization, we have a *prima facie* reason to understand and address the worry. Furthermore, insofar as we take the stakeholder theory of corporate responsibility seriously, corporations may very well have a social responsibility to reduce the extent to which employees and prospective employees experience the negative effects of hiring algorithms in their relationship with the firm. If we are right that dehumanizing the hiring algorithms will negatively impact the employee-employer relationship, then these effects are cause for concern, even if they are not decisive.

Difficult trade-offs

However, it is worth considering, and indeed even plausible, that avoiding the ills of dehumanizing the hiring process will always come at the expense of (some amount of) objectivity or impartiality. Imagine that our ballet dancer from before manages to audition before human judges rather than an algorithm. Furthermore, suppose her performance entirely captivates her human audience, some of whom are moved to tears over the nearly preternatural grace of her rendition of Odette's dance of death from Swan Lake's fourth act.

---

[7] Relatedly, Elizabeth Anderson (2017) famously defends the view that most American workers are subject to an illiberal "private government" in the workplace.

[8] See: https://www.apa.org/news/press/releases/2014/04/employee-distrust. Similar results were obtained by the American Psychological Association's more recent 2017 survey on employee well-being. See: https://www.apaexcellence.org/assets/general/2017-work-and-wellbeing-survey-results.pdf.

Our dancer is hired. The judges perhaps cannot put their fingers on exactly why her performance was so magnificent, but it likely has something to do with how each detail came together as a unique whole, unable to be reduced to specificities. The judges were moved, and they know this dance will move their audiences as well. Yet, the capacity of humans to be *moved* by other humans is the same capacity that allows bias and prejudice to enter in. Suppose our dancer had performed an identical dance for a different set of judges—judges who had each had a recent bad experience with someone who looked vaguely like our dancer. This historical fact works an emotional effect on the judges; they are, unbeknownst to themselves, less open to being moved by her dance, and so her performance leaves them feeling cold. The dancer is not hired.

Importantly, such concerns may be understood as legitimate worries about *humanization* in the hiring process. It is the *human* element, not the machine element per se, which internalizes bias and prejudice toward women, ethnic minorities, queer or gender non-conforming individuals, religious minorities, or other marginalized groups. So, it is at least worth considering whether efforts at avoiding bias in the hiring process must necessarily come at the expense of the "human element" of the relationship. We should be open to the possibility that there is a clash of values ideals. Our goal in this paper is to elucidate concerns about dehumanizing the hiring process, which—once we disambiguate this sense of dehumanization (i.e. removing the human presence) from the more common sense of dehumanization (i.e. conceiving of others as subhuman)—is not obviously morally wrong. However, it is quite possible that minimizing bias may, at times, require forfeiting some of the "humanity" of the hiring process and thus negatively impacting the employee-employer relationship. Likewise, it is probably the case that working to re-humanize the process would increase the risk of bias and prejudice. Although we lack the space to fully address this concern here, it is plausible that such trade-offs need to be decided on a case-by-case (company-by-company, job-by-job, etc.) basis. Perhaps, in the end, there is no general principle for perfectly balancing these conflicting ideals (e.g. Ross, 1930).

Conclusion

The use of AI-based tools for hiring new employees is increasing in popularity, despite some unease among HR managers. While many discussions in the ethics of algorithms are concerned with problems of bias and discrimination, in this paper we have sought to examine the issue of recruitment algorithms with an eye toward the under-explored concerns of these HR managers—in particular, the concern that the use of such technologies will "dehumanize" the hiring process. As we have argued, this sense of dehumanization, which simply involves "leaving out the human", should be carefully distinguished from the more common sense of the term, as it features in dehumanization studies, according to which dehumanization involves conceiving of others as subhuman. Clearly, nothing of the sort happens in the case of recruitment algorithms. Still, as we have argued, there is something ethically suspect about removing the human presence from the hiring process. Our argument for being suspicious of the use of hiring algorithms is that they are unable to truly take the place of a human recruitment officer, as these algorithms rely on *artificial values* that only approximate but never replicate real human values. In human relationships, our values and the relations between our values are often complex and difficult to balance. Any attempt to impose artificial value clarity on this cascade of values by means of predictive algorithms—even ones that have been subjected to rigorous validation methods—is bound to strike some employees and employers as uncomfortably alienating, thereby negatively impacting the employee-employer relationship. This concern is something that HR managers must reckon with as the presence of AI in the workplace increases in the coming years.

Declarations

References

Aglietta, M., & Rebérioux, A. (2005). *Corporate governance adrift: A critique of shareholder value*. Edward Elgar Publishing.

Anderson, E. (2017). *Private government: How employers rule our lives (and why we don't talk about it)*. Princeton University Press.

Borowski, P. (1998). Manager-employee relationships: Guided by Kant's categorical imperative or by Dilbert's business principle. *Journal of Business Ethics, 17*(15), 1623–1632.

Dastin, J. (2018). "Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women," Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

Freeman, E. (1984). *Strategic Management: A Stakeholder Approach*. Cambridge University Press.

Freeman, R. E., Harrison, J. S., & Zyglidopoulos, S. (2018). *Stakeholder Theory: Concepts and Strategies*. Cambridge University Press.

Friedman, M. (1970). "A Friedman doctrine—The Social Responsibility of Business Is to Increase Its Profits," The New York Times. https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html.

Habermas, J. (1970). Technology and Science as 'Ideology.' In J. Shapiro (Ed.), *Toward a Rational Society*. Beacon Press.

Horkheimer, M., & Adorno, T. (1972). *Dialectic of Enlightenment*. Stanford University Press.

Joerden, J. C. (2018). Dehumanization: The Ethical Perspective von Heinegg. In W. Heintschel, R. Frau, & T. Singer (Eds.), *Dehumanization of Warfare* (pp. 55–73). Springer.

Karnes, R. (2009). A change in business ethics: The impact on employer-employee relations. *Journal of Business Ethics, 87*(2), 189–197.

Kittay, E. (2005). At the margins of moral personhood. *Ethics, 116*(1), 100–131.

Kleinberg, J., Ludwig, S., Mullainathan, C., & Sunstein, C. (2020). Algorithms as discrimination detectors. *Proceedings of the National Academy of Sciences, 117*(48), 30096–30100.

Kronfeldner, M. (2021). Mapping Dehumanization Studies. In M. Kronfeldner (Ed.), *Routledge Handbook of Dehumanization* (pp. 1–36). Routledge.

Marcuse, H. (1941). Some social implications of modern technology. *Zeitschrift Für Sozialforschung, 9*(3), 414–439.

Mill, J. S. (1864). *Utilitarianism*. Cambridge University Press.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21.

Mowshowitz, A. (2008). Technology as excuse for questionable ethics. *AI & Society, 22*(3), 271–282.

Nguyen, C. T. (2021). How Twitter Gamifies Communication. In J. Lackey (Ed.), *Applied Epistemology* (pp. 410–436). Oxford University Press.

Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

Rachels, J. (1975). Why privacy is important. *Philosophy & Public Affairs, 4*(4), 323–333.

Raghavan, M., Barocas, S., Kleinberg, J., Levy, K. (2019). "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices," ACM Conference on Fairness, Accountability, and Transparency. https://arxiv.org/pdf/1906.09208.pdf.

Ross, W. D. (1930). *The right and the good*. Oxford University Press.

Schecter, D. (2010). *The Critique of Instrumental Reason from Weber to Habermas*. Continuum.

Sherman, N. (2021). "Amazon Fight with Workers: 'You're a Cog in the System'," BBC News. https://www.bbc.com/news/business-55927024

Simms, M. (1994). Defining privacy in employee health screening cases: Ethical ramifications concerning the employee/employer relationship. *Journal of Business Ethics, 13*(5), 315–325.

Smith, D. L. (2014). Dehumanization, essentialism, and moral psychology. *Philosophy Compass, 9*(11), 814–824.

Smith, D. L. (2016). Paradoxes of dehumanization. *Social Theory and Practice, 42*(2), 416–443.

Smith, D. L. (2018). Manufacturing monsters: Dehumanization and public policy. *The Palgrave Handbook of Philosophy and Public Policy* (pp. 263–275). Palgrave Macmillan.

Stout, L. A. (2012). *The Shareholder Value Myth: How Putting Shareholders First Harms Investors, Corporations, and the Public*. Berrett-Koehler Publishers.

Superson, A. (1983). The employer-employee relationship and the right to know. *Business and Professional Ethics Journal, 3*(1), 45–58.

Williams, B. (1985). *Ethics and the Limits of Philosophy*. Harvard University Press.