



From human resources to human rights: Impact assessments for hiring algorithms

Josephine Yam¹ · Joshua August Skorburg¹

Accepted: 1 June 2021 / Published online: 25 June 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

Over the years, companies have adopted hiring algorithms because they promise wider job candidate pools, lower recruitment costs and less human bias. Despite these promises, they also bring perils. Using them can inflict unintentional harms on individual human rights. These include the five human rights to work, equality and nondiscrimination, privacy, free expression and free association. Despite the human rights harms of hiring algorithms, the AI ethics literature has predominantly focused on abstract ethical principles. This is problematic for two reasons. First, AI principles have been criticized for being vague and not actionable. Second, the use of vague ethical principles to discuss algorithmic risks does not provide any accountability. This lack of accountability creates an algorithmic accountability gap. Closing this gap is crucial because, without accountability, the use of hiring algorithms can lead to discrimination and unequal access to employment opportunities. This paper makes two contributions to the AI ethics literature. First, it frames the ethical risks of hiring algorithms using international human rights law as a universal standard for determining algorithmic accountability. Second, it evaluates four types of algorithmic impact assessments in terms of how effectively they address the five human rights of job applicants implicated in hiring algorithms. It determines which of the assessments can help companies audit their hiring algorithms and close the algorithmic accountability gap.

Keywords Artificial Intelligence · Machine Learning · Human Resources · Impact Assessments · Algorithmic Audits

Introduction

It is increasingly platitudinous to remark on how Artificial Intelligence (AI) and Machine Learning (ML) applications have penetrated most sectors of the economy and are rendering highly consequential decisions that impact human lives. Still, the stakes are uniquely high for the use of AI in Human Resources (HR) recruitment. Hiring algorithms are powerful. They can determine who gets access to the economic opportunities the world offers. Particularly, they affect a person's life prospects, including what work they will do, how much they will get paid, where they will live, and the quality of life they will enjoy. In turn, these significantly impact a person's sense of autonomy, self-identity and active membership in society (Raghavan et al., 2020). The

stakes are especially high because equal access to economic opportunities is necessary for social justice (Arneson, 2015).

Over the years, companies around the world have adopted various hiring algorithms in order to win the war on talent (Bogen & Rieke, 2018). This is because they promise companies wider job candidate pools, lower recruitment costs and less human bias. Despite these promises, they also bring some perils. Using them can inflict unintentional harms on individual human rights (Amnesty International, 2019). These include the five human rights to work, equality and non-discrimination, privacy, free expression and free association (henceforth, "Five Human Rights").

Despite the human rights harms of hiring algorithms, the AI ethics literature has predominantly focused on abstract ethical principles.¹ This is problematic for two reasons.

First, it is widely recognized that "the AI industry lacks proven methods to translate principles into practice and AI principles have been criticized for being vague" (Raji et al., 2020, p. 34). For instance, the "commitments to ensure AI is 'fair', or respects 'human dignity', or enables 'human

✉ Joshua August Skorburg
skorburg@uoguelph.ca

¹ Department of Philosophy, University of Guelph, 50 Stone Rd E, Guelph, ON N1G 2W1, Canada

¹ Mittelstadt (2019) documents more than 60 such initiatives.

flourishing’ are not specific enough to be action-guiding” (Mittelstadt, 2019, p. 5).

Second, the use of vague ethical principles to discuss algorithmic risks “provid[es] little to no means of accountability” (Raji et al., 2020, p. 34). This lack of accountability arising from amorphous ethical principles creates an algorithmic accountability gap. Closing this gap is crucial because, as Raub (2018) emphasizes, “Without accountability and responsibility, the use of algorithms and artificial intelligence leads to discrimination and unequal access to employment opportunities” which violate social justice (p. 530).

This paper makes two contributions to the AI ethics literature. First, it frames the ethical risks of hiring algorithms using international human rights law as a universal standard for determining algorithmic accountability. Second, it evaluates four types of Algorithmic Impact Assessments (AIA) in terms of how effectively they address the Five Human Rights of job applicants that are implicated in hiring algorithms. It determines which of the AIAs can help companies audit their hiring algorithms and close the algorithmic accountability gap.

We have divided this paper into six sections. “Types of hiring algorithms” provides an overview of various hiring algorithms. “Benefits of hiring algorithms” and “Limitations of hiring algorithms” highlight their benefits and limitations, respectively. “Harms of hiring algorithms” establishes the Five Human Rights of job applicants that are implicated when companies use hiring algorithms. In “Overview of algorithmic impact assessments”, we consider four types of AIAs (Privacy Impact Assessment, Data Protection Impact Assessment, and Ethical and Social Impact Assessment) and evaluate them to determine which of them effectively addresses the Five Human Rights of job applicants to help close the algorithmic accountability gap. In Sect. 6, we discuss the Human Rights Impact Assessment and argue that it is best suited to close the algorithmic accountability gap when companies deploy hiring algorithms.

Types of hiring algorithms

Companies are leading the burgeoning global AI revolution. Global corporate AI spending is forecasted to increase rapidly from US \$50.1 billion in 2020 to more than \$110 billion in 2024. The use of AI-based HR tools has been identified as one of the fastest growth areas for corporate AI spending (IDC, 2020). Corporate HR departments are now using

hiring algorithms to recruit top talent at a velocity and scale never seen before.² After all, hiring top talent is crucial for achieving corporate innovation, creativity and a competitive advantage (Bogen & Rieke, 2018).

Various hiring algorithms are deployed across the four phases of the talent recruitment process (Bogen & Rieke, 2018). The first phase is sourcing, in which a company reaches out to identify and attract potential candidates to apply. For example, Textio composes job descriptions to ensure that they reach diverse job applicants using “gender tone meters” and “linguistic patterns” of past job postings relative to successful hires (Bogen & Rieke, 2018, p.15).³ Entelo scours social media to make predictions about whether possible job candidates are likely to move jobs and if they will fit a company’s culture (Bogen & Rieke, 2018). Arya spots ideal candidates from their online behavioural patterns and automatically composes personalized recruiter messages to invite them to apply (Raub, 2018).⁴

The second phase is screening, in which a company evaluates and ranks applicants based on skills, experience and personality. Highly ranked applicants are shortlisted for further screening while the rest are immediately rejected (Bogen & Rieke, 2018). For example, the chatbot Mya evaluates an applicant’s replies and sends those who it predicts will be a good fit to the interview phase.⁵ Koru predicts a candidate’s character traits through various self-assessment surveys that rank a candidate’s responses for “personal attributes like ‘grit’, ‘rigor’ and ‘teamwork’” (p. 30).⁶

Footnote 2 (continued)

rent employees to drive employee engagement and well-being. See Ajunwa, Crawford, & Ford (2016) for an overview.

³ Textio claims that its “gender tone meter” identifies overly masculine tones job postings. Its website demo explains how it works using a typical job posting which reads: “We focus on customers, collaboration, and excellence.” The word “excellence” is then flagged as a “fixed mindset phrase which emphasizes raw talent over growth and will hurt inclusion efforts”. Description retrieved from: <https://textio.com/products/> on 20 January 2021.

⁴ Arya claims to harness data from social media sites and “50+ other relevant professional sources to create meaningful insights and a deeper intelligence of each candidate” (Raub, 2018).

⁵ Mya, a “conversational AI” promises to “make hiring more human”. It requests applicants to provide additional qualification information which it independently evaluates. It then automatically schedules qualified applicants for a phone screening interview with a human recruiter. Descriptions retrieved from: <https://www.mya.com/meetmya/> on 20 January 2021.

⁶ The Cappfinity platform offers the “Koru7 Impact Skills” which tests on “the seven soft skills that every employer is looking for in their best-fit hires: Grit, Rigour, Impact, Teamwork, Curiosity, Ownership and Polish.” After completing a 20- minute assessment, an applicant is given a “fit score” and a “Koru7 impact skills profile” which predict their future job performance. If the scores on “fit” or “generosity” or if the “Koru7 impact skills profile” meet pre-determined standards, then the applicant is invited to a formal interview. Descriptions retrieved from: <https://www.cappfinity.com/koru/> on 20 January 2021.

² Beyond recruitment, companies are also using AI-based HR tools to predict and develop effective corporate wellness initiatives for cur-

Pymetrics uses interactive games to determine a candidate's "cognitive, social, and emotional traits...such as processing speed, memory, and perseverance" (p. 32).⁷

The third phase is interviewing, in which a company queries applicants with detailed questions on skills, education, prior work experiences and behaviours. In video interviews, HireVue uses speech recognition and facial analysis algorithms to evaluate a job applicant's voice, choice of words, eye contact, mood and facial expressions.⁸ It then systematically ranks applicants against each other in comparison to past successful hires (Bogen & Rieke, 2018). HireVue also employs MindX, an interactive game that evaluates an applicant's "problem-solving, mental flexibility, learning agility, attention, creativity, and quantitative aptitude" (Chaudhary, 2018, para. 3).

The last phase is selection, in which a company determines whether or not to present an applicant with a job offer with a comprehensive compensation package. Before extending job offers, many companies use Fama to perform background checks on applicants' social media activities to "flag candidates at risk of engaging in sexual harassment, workplace violence, and other 'toxic behavior'" (Bogen & Rieke, 2018, p.39).⁹ To guide them in salary negotiations, companies also use Beqom to draft a formal job offer based on where an applicant lives and how likely they will accept it.¹⁰

Having discussed the various types of hiring algorithms, we now highlight some of the benefits they provide companies.

Benefits of Hiring Algorithms

Hiring algorithms are powerful because they "combine traditional information such as work experience and education with non-traditional data including consumer and financial data and internet browsing history...and weigh how well an applicant would fit an opening" at a speed and scale which exceed even the most experienced human recruiters (Scherer, 2017, p.1).¹¹ According to LinkedIn, 64 percent of employers use AI and data analytics as part of their aggressive talent strategy to win the talent war (LinkedIn Talent Solutions, 2018). There are at least three reasons that provide companies a legitimate business case to use hiring algorithms.

First, hiring algorithms can widen the pool of diverse job candidates through targeted advertising and matchmaking platforms (Ajunwa & Schlund, 2020). Because many companies know that a diverse talent pool contributes to innovation and a competitive advantage, they leverage the power of hiring algorithms to help them achieve a diverse and inclusive workforce (Bogen & Rieke, 2018). Hiring algorithms scour through social media platforms like LinkedIn, Facebook and Twitter to spot potential job candidates. They assess the candidates' online behaviour and persona to determine if they fit the company's minimum job qualifications and corporate culture. Once they spot qualified candidates, the company's job advertisements appear on the candidates' social media feeds (Raso et al., 2018).

Second, hiring algorithms can reduce recruitment costs by screening job applicants quicker and more thoroughly. This is significant given the high ratio of applicants to a job vacancy (Raso et al., 2018). In the U.S., companies take about six weeks and spend about \$4,000 to hire a single employee (Bogen & Rieke, 2018). Because HR departments are witnessing shrinking budgets, they need to hire top calibre talent at the quickest time and lowest cost possible before their competitors do.

Lastly, hiring algorithms strive to achieve fairness by eliminating the unconscious bias that human recruiters have by using supposedly neutral technology tools (Kim, 2016). For example, instead of using resumes to assess applicants, Pymetrics uses "neuroscience games" that will "remove biases such as classism, racism, sexism, and ageism...[which do not] account for the candidate's ethnicity, educational background, referrals, or gender" (Raub, 2018, p. 539). As Raub explains, the game's objective is to ensure that the recruitment process "doesn't preference white guys from elite schools who were on the sailing team just like the recruiter" (p. 539).

⁷ Pymetrics claims to collect "objective behavioral data that measures a job seeker's true potential." It provides 12 interactive "gamified assessments" which purport to measure attention, effort, fairness, decision making, emotion, focus, generosity, learning and risk tolerance. It then develops tailored interview questions based on these assessments. Descriptions retrieved from: <https://www.pymetrics.ai/science> on 20 January 2021.

⁸ In January 2021, as a result of a public backlash, HireVue announced that it would cease using facial analysis. See, for example, <https://www.wired.com/story/job-screening-service-halts-facial-analysis-applicants/>

⁹ Fama promises to "identify problematic behavior before it becomes an issue" by providing "background checks for the twenty-first century". It flags instances of misogyny, bigotry, racism, violence and criminal behavior based on an applicant's social media activities. Descriptions retrieved from <https://fama.io/product/> on 14 January 2021.

¹⁰ Beqom uses machine learning to "optimize compensation models and incentives plans," while also claiming to keep compensation fair and close pay gaps. Descriptions retrieved from <https://www.beqom.com/artificial-intelligence-driven-compensation> on 14 January 2021.

¹¹ As we discuss in the following section, the speed and scale at which hiring algorithms perform their tasks may give rise to potential harms in some contexts.

Despite these important benefits, we now explore the inherent limitations of hiring algorithms that contribute to possible human rights harms.

Limitations of hiring algorithms

Hiring algorithms involve high-stakes decisions that impact an individual's prospects for living well. Thus, it is important to explore some inherent limitations of hiring algorithms. Here, we consider three important limitations – power asymmetries, opacity, and bias – though this list is by no means exhaustive.

Power asymmetry

A stark power asymmetry exists between companies who design and deploy hiring algorithms and the universe of potential applicants who are subjected to them. The adage that 'Information is Power' rings true in this situation because companies know a lot about potential applicants while potential applicants know practically nothing about them. On one side are the companies who "acquire very detailed, fine-grained data about [applicants]... which they can mine to generate predictions about user traits, tastes and preferences with considerable accuracy" (Yeung, 2018, p. 6). On the other side are job applicants who subject themselves to an opaque recruitment process in which they "do not understand the complexities of the digital technologies that [employers] use...[n]or do they have equivalent access to detailed information" on the company (p. 6). Yeung warns that this power asymmetry "not only expands opportunities for potential exploitation, but may steadily erode the socio-technical foundations of moral and democratic community" (p. 6).

AI opacity

Many aspects of the job application process are opaque to job candidates, even when no hiring algorithms are involved. Nevertheless, what is potentially pernicious about hiring algorithms is their inherent opacity. This opacity or 'black-box' nature of algorithms is due to their inherently dynamic nature.¹² They autonomously feed on and learn

¹² Following Burrell (2016), we can distinguish between three forms of opacity in these kinds of cases: (1) Opacity due to technical illiteracy (e.g., hiring managers insufficiently versed in statistics and thus unable to properly interpret outputs of a predictive hiring model); (2) Opacity due to competitive advantage (e.g., companies treating their data and algorithms in proprietary personality assessments as trade secrets thus limiting a third party from assessing the internal, external and construct validity of such assessments); and (3) Opacity due to fundamental representational capacities of some machine learning models (e.g., deep neural networks used in facial analysis extract-

from continuous streams of new data (Nahmias & Perel, 2020). In doing so, they constantly calibrate their models as they deliver automated decisions on whether to accept or reject a job applicant. Because of their continuous self-calibration, hiring algorithms can become so complex that they are likely unable to "produce meaningful explanations ...in terms that are intelligible to the algorithm developers" (Yeung, 2018, p. 23).

Because the automated decisions of some hiring algorithms are also unexplainable to job applicants, they may feel diminished by their lack of agency and self-determination. As Venkatasubramanian & Alfano (2020) point out, in any unfavourable decision, "the subject of the decision can only reasonably plan their subsequent course of action if they know what it would take to receive a more favorable decision. After all, a desired or hoped-for end [i.e. being employed] can only become the target of a plan if the agent is able to select a means to that end". This form of recursive planning, they correctly note, is characteristic of "mature human agency". Thus, in order to make plans to live a good life, an agent must be "positioned in such a way that they know or reasonably expect that, were things to go wrong, they would be able to set them right again" (p. 2).¹³

AI Bias

There are at least three ways in which companies can unintentionally develop and deploy biased hiring algorithms. First is through bias in data. This occurs when companies use training datasets in which "certain groups of people [are] not available in a ratio that would be required for correct representation" (Orwat, 2020, p. 55). These algorithms detect patterns of inequality in datasets and use them to make automated decisions that perpetuate or exacerbate these inequalities at scale.

Second is through bias of programmers. When programmers design algorithms, their biases, values and priorities may become embedded in code (Joh, 2017). Their personal choices are influenced by a plethora of conscious and unconscious intuitions such as "pre-existing biases, value judgments, lack of domain expertise or lived experience, or ignorance with respect to the real consequences of certain mathematical decisions for marginalized individuals and their communities" (Robertson et al, 2020, p. 120). For

Footnote 12 (continued)

ing high-level representations incomprehensible to human semantic understanding thus undermining the ability of humans to explain why or how a given classification was reached).

¹³ Even if hiring algorithms become transparent and intelligible, companies may nevertheless be unwilling to disclose such information to protect their confidential intellectual property rights over such hiring algorithms (Katyal, 2019).

example, one scholar notes that “there are at least 21 different definitions of fairness that could be used in designing an algorithm, each of which promotes certain values and politics over others” (p. 120).

Third is through biased models. Models can be biased because they embody all the complex choices programmers have made to achieve the objective function of algorithms (Orwat, 2020). Because biased choices can become encoded into a model, decision-making processes may systemically result in possible discriminatory outcomes. Unfortunately, the hiring algorithms that companies use “run the risk of codifying inequalities while providing a veneer of objectivity” (Raghavan et al., 2020, p. 470).¹⁴

Given the inherent limitations of hiring algorithms, we now examine the ethical harms they inflict through the lens of international human rights law.

Harms of hiring algorithms

In this section, we frame the ethical risks of hiring algorithms using international human rights law because it provides a consistent, universal standard for determining algorithmic accountability. As Krishnamurthy (2018) notes: “The human rights regime provides the clarity and certainty of law. It *transforms voluntary promises of ethical behavior into mandatory requirements for compliance* with an established body of law. It lays down procedures for determining when human rights have been violated and provides individuals with redress and remedy when such violations have been established... From a human rights perspective, addressing the discriminatory impacts of the algorithm ... is not a matter of ethical discretion, but of legal obligation” (para. 12, emphasis added).

When individuals apply for a job, they exercise their human right to work as provided in Article 23 of the United Nations’ Universal Declaration on Human Rights (UDHR). In doing so, however, they subject themselves to an opaque and murky recruitment process that requires them to give

¹⁴ The infamous Amazon talent recruitment case is a classic example of AI bias. In 2015, Amazon created hiring algorithms to scout for software engineers. They used training data based on the resumes of their top software engineers over the past decade (Houser, 2019). Because 80 percent of Amazon’s top software engineers were male, the algorithms were primarily trained with their resumes. Even if the algorithms were explicitly designed not to use gender as a factor, they recognized patterns in the resumes which had “verbs more commonly found on male engineers’ resumes, such as ‘executed’ and ‘captured’” (Dastin, 2018). Similarly, Amazon’s hiring algorithms may have rejected most female applicants because their resumes contained the term ‘women’s’ due to joining a women’s tennis club or studying at an all-women’s college (Tambe et al., 2019). When Amazon discovered these discriminatory results, it stopped using such hiring algorithms immediately.

up other human rights. In this Faustian bargain, agents are only able to enjoy their human rights online by submitting to a system predicated on human rights violations (Amnesty International, 2019, p. 6).

Thus, in order to achieve algorithmic accountability, companies should identify the various human rights harms that their hiring algorithms can cause. Hiring algorithms impact not only the human right to work but also the rights to equality and non-discrimination, privacy, free expression and free association (United Nations, 2018). We refer to these rights as the “Five Human Rights” of job applicants. In the next section, we will evaluate whether various Algorithmic Impact Assessments (AIA) sufficiently address these Five Human Rights in order to close the algorithmic accountability gap.

Right to equality and non-discrimination

This right is found in Article 2 of the UDHR. It constitutes the core of human dignity pursuant to the UDHR’s explicit articulation that “all human beings are born free and equal in dignity and rights” (United Nations, 2018, p. 170). This right is implicated when patterns of systemic bias and discrimination are detected from training data and reproduced by the algorithms to create digital profiles of potential applicants (Raso et al., 2018). As discussed earlier, hiring algorithms can be skewed unfairly because of the biases of training data, programmers and models. Thus, they can powerfully reproduce and amplify habitual patterns of bias and discrimination that pervade the current employment domain. Even if specific individual protected data points such as gender or race are purposely removed from the algorithmic model, algorithms can detect silent and discreet statistical correlations amongst other data points (The Leadership Conference Education Fund (The Leadership Conference Education Fund, 2020).

Unfortunately, these types of discriminatory outcomes “produce results that are no better, and sometimes much worse, than those hewn from the ‘crooked timber of humanity’” (Raso et al., 2018, p. 7). As Kim (2016) observes, “When these automated decisions are used to control access to employment opportunities, the results may look very similar to the systematic patterns of disadvantage that motivated antidiscrimination laws. What is novel is that the discriminatory effects are data-driven” (p. 861). Moreover, it is difficult to make companies accountable for such discriminatory impacts because the cloak of confidentiality often shrouds their proprietary hiring algorithms.

Right to privacy

The right is found in Article 12 of the UDHR. It encompasses “protection of individual autonomy” as well as

“protection of personal information” (United Nations, 2018, p. 94). This right explicitly recognizes that all individuals possess “the right to achieve self-realization with full respect for their individuality and autonomy” (p. 94). An applicant’s right to privacy recognizes their personal autonomy to control the way they present themselves in their job applications and divulge only the information that they want potential employers to know (Ajunwa & Schlund, 2020).

Over the past few years, this right has been severely tested and challenged by companies who have scrambled to continuously feed their AI-based systems which vast amounts of data. Companies aggressively leverage AI’s ability to harness massive amounts of personal data from non-traditional sources such as social media and other online behaviour. The algorithms integrate such data into detailed digital profiles of individuals that employers use to make hiring and promotion decisions. These digital profiles of applicants are often relied upon by companies without applicants being aware of them or consenting to them. This potentially infringes their rights to personal autonomy and self-determination.

For example, one company used the correlation of the commute length between its office headquarters and a job candidate’s home to predict likelihood of their long employment tenure. Kim (2016) cautions that “if a hiring algorithm relied on that factor, it would likely have a racially disproportionate impact, given that discrimination has shaped residential patterns in many cities” (p. 863). Along similar lines, Yeung (2018) keenly observes, “The use of personal data for the purposes of individual profiling, and its subsequent repurposing, threatens a person’s right to ‘informational self-determination’ particularly given that ... even fairly mundane, innocuous data collected from the digital traces of individuals may ... enable quite intimate personal information to be inferred at a very high level of accuracy” (p. 37).

Right to freedom of opinion and expression

This right is found in Article 19 of the UDHR. It recognizes that the freedom to express one’s own ideas is vital in “a democratic society where there is a flow of ideas and information” (United Nations, 2018, p. 89). It includes the freedom to keep confidential or to disclose information about oneself. All individuals have the right to autonomy and self-determination such that they can control how and when they wish to present their identity to the world. Like the right to privacy, this right is implicated when hiring algorithms integrate an applicant’s entire history of online activities on social media platforms into a digital profile of a job applicant without their knowledge or consent. Because of this information asymmetry, job applicants rarely, if ever, have the opportunity to confirm, correct or challenge the digital identities that hiring algorithms automatically create, which becomes the basis of a company’s hiring decisions.

In this regard, authentic forms of self-expression can be systematically undermined in two ways: first, by limiting authentic expression owing to the fear of decreasing levels of employability; and second, by promoting inauthentic forms of expression meant to mimic desirable qualities to increase employability.¹⁵ Particularly, hiring algorithms can hinder people from expressing themselves in ways that may harm their chances of employment. As Raso et al. (2018) explain, this may harmfully impact people’s freedom of expression because it “may chill individuals from engaging in certain forms of expressive activity out of fear that their words will be used against them in the employment context” in ways they are not aware of or cannot foresee (p.46). This potentially violates an individual’s rights to freedom of self-expression, privacy and personal autonomy.

Thus, it is easy to understand why individuals would limit their forms of self-expression in social media. Anything they say, post, tweet, share or like can be used against them by potential employers. Indeed, the Council of Europe notes that, “once people know they are being surveyed, they start to behave and develop differently” (Council of Europe, 2020, p. 8). In turn, this potentially compromises an individual’s ability to engage in various forms of self-realization. When an individual exercises their right to free expression, it can become a revelation of their identity not only to their community, but also to themselves. As Gilmore (2011, p. 518) notes, “sometimes we don’t seem to know what our desires, beliefs, and commitments are until we try to express and convey them to others...[S]elf-realization is sometimes internally related to the very activity of expression. Expressing ourselves is one way in which we come to form and know our own minds”. This form of self-realization is crucial for the development of autonomy, self-improvement and self-fulfillment (Lim, 2013, p.13).

Right to freedom of association

This right is found in Article 20 of the UDHR. It recognizes that every individual has the right to voluntarily form, join, not join or leave an organization (United Nations, 2018). Hiring algorithms can have an adverse impact on this right because they can hinder people from freely associating with certain groups that may limit their job prospects. Like the right to freedom of opinion and expression, this right speaks to the tremendous capability of hiring algorithms

¹⁵ To illustrate, Hootsuite, the leading social media management platform, claims to “empower employees to share posts across their own social networks” under the heading of “employee advocacy”. However, it should be done in a way that “reduces risk of non-compliant or off-brand posts by providing only approved messages for staff to share.” Retrieved from <https://www.hootsuite.com/solutions/employee-advocacy> on 15 February 2021.

to correlate scattered pieces of online personal data that identify an applicant as a member of various groups and communities. For example, they can detect an applicant's associations with various groups by consolidating in one profile the websites visited, posts liked and tweets shared by the applicant for a human recruiter to review.

Because all data is potentially hiring data, Raso et al. (2018) explain that “people may be chilled ... from associating with certain others for fear of the impact on their employability” (p. 42). As an example of gender discrimination, “women with a leadership role in a student LGBTQ organization received 30% fewer callbacks for a job posting than applicants with an identical resume but without the LGBTQ association” (p. 44). As a result, if women find out that membership in an LGBTQ group can harm their employment opportunities, they “will feel compelled to disassociate themselves from organizations that might hurt their chances of securing employment, because AI-based systems may be more likely to detect such associations than human recruiters” (Raso et al., 2018, p. 46).

Right to work

This right is found in Article 23 of the UDHR. No one can be denied the opportunity “to gain one’s living by work that is freely chosen or accepted” (United Nations Human Rights Regional Office for Europe, 2018, p. 107). This right guarantees that everyone is entitled to “fair wages, equal remuneration for work of equal value, and the ability to afford a decent living for themselves and their families” (p. 107).

Indeed, there are many cases of hiring algorithms violating the human right to work of historically disadvantaged groups. For example, hiring algorithms that use voice analysis and facial recognition have been shown to make grossly inaccurate assessments on “people of color, English speakers with non-native accents, and transgender, nonbinary, and gender nonconforming people” (The Leadership Conference Education Fund, 2020, p. 2). In one case, a hiring algorithm determined that data points such as being named “Jared” and “playing lacrosse” were strong predictors of highly successful employees.¹⁶ Likewise, tools trained on biased historical data of successful top male employees may exacerbate existing pay inequalities where men are paid more than women for exactly the same job (Raso et al., 2018). Like the right to equality and non-discrimination, this right is implicated because of the AI opacity and AI bias inherent in some hiring algorithms.

¹⁶ This is an actual case, reported by *Quartz*, where the name “Jared” and “playing lacrosse” were identified by a hiring algorithm as top predictors of job success. See: <https://qz.com/1427621/companies-are-on-the-hook-if-their-hiring-algorithms-are-biased/>

Having identified the Five Human Rights of job applicants, we will now discuss the role of an AIA in auditing hiring algorithms. We will then evaluate the different types of AIAs to determine which of them sufficiently addresses the Five Human Rights in order to close the algorithmic accountability gap.

Overview of Algorithmic Impact Assessments

An impact assessment is “the process of identifying the future consequences of current or proposed action” (Nahmias & Perel, 2020, p.15). Companies use it to prospectively evaluate the possible positive and negative impacts of a project (International Association for Impact Assessment [IAIA], 2012). For decades, governments around the world have passed regulations that required impact assessments to evaluate the risks or effects of specific types of projects.

To close the algorithmic accountability gap, companies can conduct an AIA of their AI-based systems (Nahmias & Perel, 2020). The AIA is a corporate self-assessment mechanism that determines the risks that may arise from the design, development and deployment of AI-based projects, including hiring algorithms. Part of the AIA’s risk assessment is an evaluation of how those risks can be addressed, mitigated or eliminated in order to successfully implement the project (IAIA, 2012). Thus, impact assessments can help demystify the ‘black-box’ opacity of hiring algorithms. They may also help make the automated decisions of hiring algorithms explainable and correct any discriminatory decisions that may be rendered on individuals (Nahmias & Perel, 2020).

Three major regulatory frameworks for algorithmic accountability have emerged over the past few years. They require or will require companies to conduct some variant of AIAs to assess the impacts of their AI-driven systems (Chew et al., 2020). First is the E.U.’s General Data Protection Regulation of 2018 (“GDPR”). This landmark regulation constitutes the European Union’s overarching framework for the collection and use of individual personal data (Nahmias & Perel, 2020). It requires organizations to conduct a data protection impact assessment as an important accountability tool. Second is the proposed U.S. Algorithmic Accountability Act. Its goal is to ensure that companies regularly assess their AI-driven systems “for impacts on accuracy, fairness, bias, discrimination, privacy, and security” (Algorithmic Accountability Act, 2019). If it becomes law, companies will be required to conduct “automated decision system impact assessments and data protection impact assessments” of their AI-based systems (Chae, 2020, p. 22). Third is the Canada’s proposed Consumer Privacy Protection Act (“CPPA”). It will require companies to conduct privacy impact assessments to

ensure that their AI-based systems respect human rights to demonstrate algorithmic accountability (Office of the Privacy Commissioner of Canada, 2020).

Likewise, several U.S. sub-national jurisdictions are regulating companies for algorithmic accountability. California passed the California Privacy Rights Act of 2020 (CPRA), which requires businesses to conduct “a risk assessment with respect to their processing of personal information” (California Privacy Rights Act, 2020).

In terms of automated hiring systems, Illinois passed the Artificial Intelligence Video Interview Act which requires companies to publicly disclose if automated hiring systems will be used for video interviews (Jimenez, 2020). Likewise, New York City recently introduced a bill that mandates annual bias audits of automated hiring systems and will be under the jurisdiction of the NYC Commission on Human Rights (O’Keefe et al., 2020).

When companies audit their hiring algorithms, they should use a suitable AIA that effectively addresses the Five Human Rights of job applicants. This is because a suitable AIA will help companies close the algorithmic accountability gap. We will now evaluate the various types of AIAs to determine which of them is a suitable AIA when auditing hiring algorithms.

Privacy impact assessment

At the dawn of the information technology revolution in the 1990s, the European Commission required companies to include a privacy impact assessment (PIA) as a regulatory tool for data protection (Wright & Friedewald, 2013). In 2018, the European Commission included the PIA as part of the General Data Protection Regulation (GDPR). The PIA became the Data Protection Impact Assessment under the GDPR (Wright, 2013).

The PIA is a tool for appraising and mitigating the potential adverse consequences that a project, service or product may have on an individual’s privacy (Raab, 2020). It aims to protect the various facets of an individual’s privacy including “privacy of the person, privacy of personal behaviour and privacy of personal communications, as well as privacy of personal data” (Clarke, 2009, p. 124). It helps companies appreciate that a potential breach of individual privacy can considerably hurt their corporate brands, erode stakeholder trust and decrease shareholder value (Binns, 2017).

One major shortcoming of the PIA regulatory framework is it lacks a universal standard for privacy. Companies can use their unfettered discretion to draft their own definitions of privacy. They can freely determine their hiring algorithm’s outcomes, which they believe have minimal impact on personal data (Clarke, 2009). Lacking a universal standard to guide corporate conduct, the PIA has been reduced to a ‘tick-the-box’ exercise to meet minimum compliance.

Moreover, the PIA regulatory framework is exclusively focused on the right to privacy. Such right is only one of the Five Human Rights of job applicants. Its narrow scope fails to address the other four human rights violations that may emanate from privacy breaches. For example, it fails to address the possible infringement of the right of free expression to disclose only the personal data that a job applicant wants potential employers to know about them. Because of this, Mantelero (2018) notes, “The PIA models mainly focus on the individual dimension of data protection and ignore the ethical and social issues... leaving little room for other fundamental rights and freedoms” (p.768).

For these reasons, we contend that the PIA is not a suitable standalone AIA because it does not effectively address the Five Human Rights of job applicants. Thus, if companies solely use the PIA when auditing their hiring algorithms, it will not close the algorithmic accountability gap.

Data protection impact assessment

In 2018, the European Commission promulgated the GDPR, whose objective is “to give individuals more control over their personal data” (Nahmias & Perel, 2020). It is considered as the “global gold standard for privacy regulation” (p.19). The GDPR framework enhances the PIA framework by providing a universal standard, instead of a discretionary standard, on individual privacy. It provides individuals with a family of privacy rights such as the right to be informed, right of access, right to rectification, right to erasure, right to restrict processing, right to data portability, right to object and other rights in relation to automated decision making and profiling (ICO, 2020).

Section 35 of the GDPR introduced the Data Protection Impact Assessment (DPIA) as an important accountability tool. The DPIA ensures that regulated organizations comply with their data protection obligations, which entail the “fair and proper use of information about people, [which is] part of the fundamental right to privacy” (ICO, 2020, DPIA section). It helps organizations “systematically analyse, identify and minimise the data protection risks of a project” including an automated hiring system (ICO, 2020, DPIA section).

While the DPIA is primarily fixated on the rights to data protection and privacy, it may deal with “other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion” (Raab, 2020, p. 8). Nevertheless, one scholar cautions that, despite the GDPR’s explicit references to human rights, the DPIA practically remains a process-heavy exercise that concentrates on data governance. It does not provide substantive guardrails as to how companies should respect human rights (Mantelero, 2018).

Like the PIA, the DPIA is primarily focused on the right to privacy. Such right is only one of the Five Human Rights

of job applicants. The GDPR framework does not encompass the other four fundamental rights impacted by hiring algorithms. Because of the DPIA's narrow scope on personal data protection, it cannot meaningfully address the other possible pernicious human impacts of automated hiring systems.

For these reasons, we contend that the DPIA is not a suitable standalone AIA because it does not effectively address the Five Human Rights of job applicants. Thus, if companies solely use the DPIA when auditing their hiring algorithms, it will not close the algorithmic accountability gap.

Ethical and social impact assessment

The Ethical and Social Impact Assessment (ESIA) is a combination of the Ethical Impact Assessment (EIA) and the Social Impact Assessment (SIA). Despite their different origins, the EIA and SIA share a common goal of identifying, understanding and managing all the positive and negative outcomes related to well-being (Research Centre of the Slovenian Academy of Sciences Arts, 2017). Companies who carry out an ESIA do so as part of “promulgating and voluntarily committing themselves to abide by so-called ‘ethical standards’” (Yeung, 2018, p. 7). To date, the ESIA has gained some prominence as an AIA. This is because of its all-embracing focus on human well-being, which is the lens used to appreciate the expansive nature of ethical and social impacts of AI-based systems.

Despite its expansive scope, the ESIA lacks a universal standard of ethical principles that companies can benchmark against when carrying out an ESIA. Inevitably, a vast array of ethical guidelines, principles, charters and codes of conduct have emerged that companies can cherry-pick from when they conduct an ESIA of their hiring algorithms (Esteves et al., 2017, p. 76).

In fact, by Mittelstadt (2019)'s count, “at least 63 such ‘AI Ethics’ initiatives have published reports describing high-level ethical principles, tenets, values, or other abstract requirements for AI development and deployment” (p. 1). Indeed, the massive volume of ethical AI frameworks has created a “problem of ‘principle proliferation’ which threatens to overwhelm and confuse” not only companies and their users but even society at large, observe Floridi and Cowls (2019, p. 2). They note that 47 different ethical AI principles exist resulting in unnecessary duplication, overlap and confusion among various stakeholders. The phenomenon of “principle proliferation” ironically facilitates “a ‘market for principles’ where stakeholders may be tempted to ‘shop’ for the most appealing ones” in their mad race to the bottom of what is truly ethical (p. 3).

To illustrate, HireVue released an algorithmic audit report of one of its hiring algorithms in January 2021. The report was based on the ESIA conducted by Cathy

O’Neil of ORCAA, an AI auditing firm (Zuloaga, 2021). The scope of the ESIA focused on the ethical principles of fairness and bias relating to HireVue’s pre-built assessments used for recruiting early-career applicants. The report concluded that the assessments “work as advertised with regard to fairness and bias issues” (ORCAA, 2020, p. 3). Interestingly, the report qualified its conclusion, noting that “for the issues of fairness and bias ...there are few hard and fast rules...There will be grey areas and ethical dilemmas; it comes with the territory” (O’Neil Risk Consulting and Algorithmic Auditing (ORCAA), 2020, p. 7). In a subsequent interview, ORCAA’s chief strategist admitted the inherent mutable standards in an ESIA-based algorithmic audit. He explained that “*an algorithm audit is currently defined as whatever a selected algorithm auditor is offering...*[thus] the need for an environment where there are legal and regulatory requirements that give some more momentum to the auditors” (Johnson, 2021, emphasis added).

Thus, there is a real risk that companies that use the ESIA can employ whatever fairness definition suits their interests. This can result in companies concealing, rather than addressing, possible pernicious forms of algorithmic bias in their hiring algorithms. As Schellmann (2021) observes, “at best, audits give an incomplete picture, and at worst, they could help companies hide problematic or controversial practices behind an auditor’s stamp of approval.”

It is important to note, however, that an ESIA explicitly covers more ground than the PIA and DPIA in addressing the Five Human Rights of job applicants. It treats both privacy and data protection as only two of many other ethical principles it applies when assessing hiring algorithms (Wright & Friedewald, 2013). Nevertheless, the ESIA lacks a universal benchmark of ethical principles that companies can use when carrying out an ESIA, such as fairness, equity and human dignity. Mittelstadt (2019) sums up the situation: “We must... hesitate to celebrate consensus around high-level principles that hide deep political and normative disagreement. Shared principles are not enough to guarantee ‘Trustworthy AI’ or ‘Ethical AI’ in the future” (p. 9).

For these reasons, we contend that the ESIA is not a suitable standalone AIA because it does not effectively address the Five Human Rights of job applicants. Thus, if companies solely use the ESIA when auditing their hiring algorithms, it will not close the algorithmic accountability gap.

Having reviewed the PIA, DPIA and ESIA, we determined that none of them effectively addresses the Five Human Rights of job applicants that are impacted by hiring algorithms. Thus, none of them is a suitable standalone AIA to help companies audit their hiring algorithms and close the algorithmic accountability gap.

Human rights impact assessment

In this Section, we evaluate the Human Rights Impact Assessment (HRIA) to determine if it is a suitable standalone AIA when companies audit their hiring algorithms to help close the algorithmic accountability gap.

In 2011, the Human Rights Council approved the United Nations Guiding Principles on Business and Human Rights (“UNGP”). The United Nations (2012) proclaimed it as the “global standard of practice” that brings “clarity about the baseline expectations of business” with regard to human rights (p. 1). The HRIA came into being when the UNGP declared that businesses have a responsibility to respect human rights by conducting human rights due diligence of their operations (Gotzmann, 2017). While the UNGP did not prescribe a specific assessment tool for this purpose, many businesses use the HRIA as a tool to help them fulfill their due diligence responsibilities.

The HRIA provides businesses with a systematic methodology to identify, assess, manage, mitigate and remedy the adverse human impacts their activities cause on individuals and communities (Gotzmann, 2017). In essence, it asks businesses to meaningfully respond to the question “How does the project, policy or intervention affect human rights?” (p.15). At its very core, the HRIA encompasses a human rights lens to promote and protect human dignity. As McGregor et al. (2019) pronounces, using a human rights lens for AIAs “offers an organizing framework for the design, development and deployment of algorithms, and identifies the factors that States and businesses should take into consideration in order to avoid undermining, or violating, human rights” (p. 313). Thus, when companies use the HRIA for their automated hiring systems, they are able to “identify, anticipate, and minimize an important class of risks and harms... so that the societal impact of AI systems [are viewed] through the lens of human rights” notes Latonero (2018, p. 6). He sums up the HRIA’s overall advantage as a powerful AIA tool in this way: “[The HRIA] offers a strong value proposition: an approach to AI governance that upholds human dignity based on international human rights law” (p. 7).

The HRIA framework offers international human rights law as its universal standard that encompasses the Five Human Rights of job applicants (Gotzmann et al., 2016). It enables companies and regulators to evaluate HRIAs in a systematic and meaningful manner by providing a consistent set of standards. These standards ordain the HRIA with legitimacy so that its findings are deemed credible, trusted and reliable (Gotzmann et al., 2016). The key strength of the HRIA is that “human rights, as a language and legal framework, is itself a source of power because human rights carry significant moral legitimacy and the

reputational cost of being perceived as a human rights violator can be very high” (Latonero, 2018, p. 9).

The universal standards of the HRIA are overarching and comprehensive yet also aim to be sensitive to culture nuances. When States ratified the UDHR, they committed to upholding those human rights by adapting and contextualizing them through appropriate instruments including national legislation. In this regard, companies have a shared understanding of the goals, benefits and harmful risks of their hiring algorithms. For example, the right to equality and non-discrimination is a universal standard found in Article 2 of the UDHR. In the U.S., this right is contextualized and made enforceable in Title VII of the Civil Rights Act of 1991, a comprehensive anti-discrimination law in the employment context (Kim, 2016). Thus, when companies want to assess whether or not their hiring algorithms amount to wrongful discrimination, they can turn to the law, which attempts to articulate what acts amount to lawful as opposed to wrongful discrimination (Nahmias & Perel, 2020).

Consequently, the meaning of wrongful discrimination will not be held hostage to a company’s subjective interpretation because the law is there to provide guidance and enforce compliance. With a shared understanding provided by national legislation, companies will have a “common, comparable threshold” that they can use to benchmark themselves against in their sector for purposes of corporate accountability for human rights (Nahmias & Perel, 2020, p. 42).

McGregor et al. (2019) eloquently explain the significance of universal standards in determining ‘harm’ in algorithmic accountability discourse:

Within the literature on algorithmic accountability, the term ‘bias’ (and less often ‘discrimination’) is used in a range of different ways, often without clarity on the meaning employed. It is sometimes used to convey a specific technical meaning, for example with reference to statistical bias. In other contexts, it is employed as a general, ‘catch-all’ term to mean some form of preference or ‘unfairness’ (which itself has been criticized as a vague term). When used in such a broad way, actors may develop or gravitate to locally defined understandings as to what constitutes bias or discrimination, giving rise to a variety of meanings. This can also create uncertainty for actors designing, developing and using algorithms in decision-making as to whether a particular instance of bias is unlawful.

[International Human Rights Law or “IHRL”] provides a concrete and universally applicable definition of harm that is capable of identifying prohibited and unlawful forms of bias and discrimination. This definition is accompanied by well-developed and sophisticated tests for establishing when the prohibition

of discrimination has been violated, including what constitutes direct, indirect or intersectional discrimination as well as structural and unconscious bias. IHRL therefore not only provides a means to determine harm through its interpretation of how rights may be interfered with, it also provides established tests to assess when and how rights may have been violated. (p.325-6)

For these reasons, we argue that the HRIA is a suitable standalone AIA because it effectively addresses the Five Human Rights of job applicants. This is in stark contrast to the PIA and DPIA which are focused on only one of the Five Human Rights, the right to privacy. This is also in stark contrast to the ESIA which does not provide a universal benchmark to address the Five Human Rights and is “not based on legal standards” (Gotzmann et al., 2020, p. 25). In this regard, if companies use the HRIA when auditing their hiring algorithms, it will help close the algorithmic accountability gap.

Conclusion

Over the years, companies have increasingly adopted hiring algorithms to help them win the talent war. Despite their invaluable benefits, hiring algorithms can inflict significant harms to individual human rights. Because of issues such as opacity and bias, hiring algorithms can negatively impact an individual’s Five Human Rights to work, equality and non-discrimination, privacy, free expression and association. The inability to exercise these rights has consequential impacts on a person’s sense of autonomy, self-identity and active membership in society. Because millions of people scout for new jobs annually, hiring algorithms potentially “enable human rights harm at a population scale” (Amnesty International, 2019, p. 6).

Unless national legislation is put in place, companies may not be legally accountable for the human rights infringements of their hiring algorithms. This creates an algorithmic accountability gap. One of the ways companies can address this gap is to proactively conduct an AIA of their hiring algorithms.

Depending on the context in which the HRIA is conducted, using the HRIA as a standalone impact assessment may not always be enough to address the myriad risks and untold effects that algorithms may have. In a recent paper, Metcalf et al. (2021) cite how difficult it is to ensure that the kinds of impacts identified within the scope of AIAs are accurately related to actual or potential harms to those most likely to experience them. As they point out, “What counts as an adequate assessment, when that assessment happens, and how stakeholders are made accountable to each other

are contested outcomes shaped by fraught power relationships” (Metcalf et al., 2021 p. 1). With hiring algorithms, the power asymmetry between companies who deploy them and the universe of potential applicants who are subjected to them will be especially salient. In this regard, companies should be at liberty to use the HRIA in combination with the other AIAs. This combined approach will facilitate comprehensive stakeholder consultations to understand the full range of impacts involved. Thus, non-human rights impacts that fall outside the HRIA’s scope can likewise be squarely addressed.

When companies use the HRIA in auditing their hiring algorithms, they can aim to ensure the resulting impacts are rights-respecting, if not rights-enhancing. That means that companies should give every job applicant an equal opportunity to be fairly and justly considered before a decision is made to accept or reject their application. This may require a human recruiter-in-the-loop at all times during the four phases of the algorithmic hiring process. After all, only humans, not hiring algorithms, will know how to treat fellow humans with fairness, empathy, dignity and respect. Especially when an outright rejection of their application really means ‘Sorry but you’re just not good enough to work here’.

Thus, the HRIA is an invaluable tool that companies should always have in their self-regulatory and compliance toolbox. Because it is a suitable standalone AIA that addresses the Five Human Rights of job applicants, companies can capitalize on its robustness to right the wrongs of hiring algorithms (Nahmias & Perel, 2020). Consequently, when engaging in the talent war, companies should use the HRIA to ensure that their powerful hiring algorithms safeguard human rights, not abuse them.¹⁷

Funding None.

Declaration

Disclosure The authors report no conflict of interest.

References

- Ajunwa, I., & Schlund, R. (2020). Algorithms and the social organization of work. *The Oxford Handbook of Ethics of AI*. <https://doi.org/10.1093/oxfordhb/9780190067397.013.52>
- Ajunwa, I., Crawford, K., & Ford, J. S. (2016). Health and big data: An ethical framework for health information collection by corporate wellness programs. *The Journal of Law, Medicine & Ethics*, 44(3), 474–480.

¹⁷ Thanks to Don Dedrick for helpful feedback on an early draft of this manuscript.

- Algorithmic Accountability Act of 2019, S. 1108, H.R. 2231, 116th Cong. (2019) <https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info>.
- Amnesty International. (2019). *Surveillance Giants: How The Business Model of Google and Facebook Threatens Human Rights* (p. 60). Amnesty International.
- Arneson, R. (2015). Equality of Opportunity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2015)*. California: Metaphysics Research Lab, Stanford University.
- Binns, R. (2017). Data protection impact assessments: A meta-regulatory approach. *International Data Privacy Law*, 7(1), 22–35. <https://doi.org/10.1093/idpl/1pw027>
- Bogen, M., & Rieke, A. (2018). *Help Wanted—An Exploration of Hiring Algorithms, Equity and Bias*. (p. 75). Upturn. <https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20-%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf>.
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512.
- California Privacy Rights Act of 2020 (2020). https://iapp.org/media/pdf/resource_center/ca_privacy_rights_act_2020_ballot_initiative.pdf.
- Chae, Y. (2020). U.S. AI regulation guide: legislative overview and practical considerations. *The Journal of Robotics, Artificial Intelligence & Law*, 3(1), 17–40.
- Chaudhary, M. (2018, May 14). *HireVue Acquires MindX to Create a Robust AI-Based Talent Assessment Suite*. <https://www.hrtechnologist.com/news/recruitment-onboarding/hirevue-acquires-mindx-to-create-a-robust-ai-based-talent-assessment-suite/>.
- Chew, B., Rae, J., Manstorf, J., & Degnegaard, S. (2020). *Government Trends 2020: What are the most transformational trends in government today?* (p. 88) [Deloitte Center for Government Insights]. Deloitte Consulting LLP. <https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/public-sector/lu-government-trends-2020.pdf>.
- Clarke, R. (2009). Privacy impact assessment: Its origins and development. *Computer Law & Security Review*, 25(2), 123–135. <https://doi.org/10.1016/j.clsr.2009.02.002>
- Council of Europe. “Ad Hoc Committee On Artificial Intelligence (CAHAI) - Feasibility Study,” 2020. <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-1680a0c6da>.
- Dastin, J. (2018, October 11). Insight—Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. <https://in.reuters.com/article/amazon-com-jobs-automation-idINKCN1MK0AH>
- Esteves, A. M., Factor, G., Vanclay, F., Götzmann, N., & Moreira, S. (2017). Adapting social impact assessment to address a project’s human rights impacts and risks. *Environmental Impact Assessment Review*, 67, 73–87. <https://doi.org/10.1016/j.eiar.2017.07.001>
- Floridi, L., & Cows, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.8cd550d1>
- Gilmore, J. (2011). Expression as realization: speakers’ interests in freedom of speech. *Law and Philosophy*, 30(5), 517–539.
- Gotzmann, N. (2017). Human rights impact assessment of business activities: key criteria for establishing a meaningful practice. *Business and Human Rights Journal*, 2(1), 87–108. <https://doi.org/10.1017/bhj.2016.24>
- Gotzmann, N., Vanclay, F., & Seier, F. (2016). Social and human rights impact assessments: What can they learn from each other? *Impact Assessment and Project Appraisal*, 34(1), 14–23. <https://doi.org/10.1080/14615517.2015.1096036>
- Gotzmann, N., Bansal, T., Wrzoncki, E., Veiberg, C. B., Tedaldi, J., & Høvsgaard, R. (2020). *Human rights impact assessment guidance and toolbox* | *The Danish Institute for Human Rights*. The Danish Institute for Human Rights. <https://www.humanrights.dk/business/tools/human-rights-impact-assessment-guidance-toolbox>.
- Houser, K. (2019). *Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making*. 65.
- International Data Corporation. (2020, August 25). *Worldwide Spending on Artificial Intelligence Is Expected to Double in Four Years, Reaching \$110 Billion in 2024, According to New IDC Spending Guide*. IDC: The Premier Global Market Intelligence Company. <https://www.idc.com/getdoc.jsp?containerId=prUS46794720>.
- Information Commissioner’s Office. (2020, July 20). *Data protection impact assessments*. ICO - Guide to the General Data Protection Regulation (GDPR), Accountability and Governance; ICO. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/>.
- International Association for Impact Assessment (IAIA). (2012). *Fastaps No. 1 Impact Assessment*. IAIA.
- Joh, E. (2017). Feeding the machine: policing, crime data, & algorithms symposium: Big data, national security, and the fourth amendment. *William & Mary Bill of Rights Journal*, 26(2), 287–302.
- Johnson, K. (2021). What algorithm auditing startups need to succeed. *VentureBeat*. <https://venturebeat.com/2021/01/30/what-algorithm-auditing-startups-need-to-succeed/>.
- Katyal, S. K. (2019). Private Accountability in the Age of Artificial Intelligence. *UCLA Law Review*, 66(1), 54–141.
- Khan, A. N., Ihalage, A. A., Ma, Y., Liu, B., Liu, Y., & Hao, Y. (2021). Deep learning framework for subject-independent emotion detection using wireless signals. *PLoS ONE*, 16(2), e0242946.
- Kim, P. T. (2016). Data-driven discrimination at work. *William & Mary Law Review*, 58(3), 857–936.
- Krishnamurthy, V. (2018, October 10). It’s not enough for AI to be “ethical”; it must also be “rights respecting.” *Berkman Klein Center for Internet & Society at Harvard University*. <https://medium.com/berkman-klein-center/its-not-enough-for-ai-to-be-ethical-it-must-also-be-rights-respecting-b87f7e215b97>.
- Kroll, J. (2020). Accountability in computer systems. *The Oxford Handbook of Ethics of AI*. <https://doi.org/10.1093/oxfordhb/9780190067397.013.10>
- Latonero, M. (2018). *Governing Artificial Intelligence: Upholding Human Rights & Dignity* (p. 38). Data & Society. <https://datasociety.net/library/governing-artificial-intelligence/>.
- Lim, M. (2013). Freedom of expression toolkit: A guide for students. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000218618>.
- LinkedIn Talent Solutions. *LinkedIn Global Recruiting Trends 2018*. (2018). LinkedIn Talent Solutions. <https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/resources/pdfs/linkedin-global-recruiting-trends-2018-en-us2.pdf>.
- Mantelero, A. (2018). AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review*, 34(4), 754–772. <https://doi.org/10.1016/j.clsr.2018.05.017>
- McGregor, L., Murray, D., & Ng, V. (2019). International human rights law as a framework for algorithmic accountability. *International & Comparative Law Quarterly*, 68(2), 309–343. <https://doi.org/10.1017/S0020589319000046>
- Metcalfe, J., Moss, E., Watkins, E. A., Singh, R., & Elish, M. C. (2021). *Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts*. 19. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3736261.
- Mittelstadt, B. (2019). AI ethics—too principled to fail? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3391293>

- Nahmias, Y., & Perel, M. (2020). The Oversight of Content Moderation by AI: Impact Assessments and Their Limitations. *Harvard Journal on Legislation*, 54. <https://papers.ssrn.com/abstract=3565025>.
- Office of the Privacy Commissioner of Canada. (2020, November 12). *A Regulatory Framework for AI: Recommendations for PIPEDA Reform*. https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai-reg-fw_202011/.
- O’Keefe, J., Moss, D. J., & Martinez, T. S. (2020, March 10). Mandatory “Bias Audits” and Special Notices to Job Candidates: New York City Aims to Regulate the Use of Artificial Intelligence in the Workplace. *Law and the Workplace*. <https://www.lawandtheworkplace.com/2020/03/mandatory-bias-audits-and-special-notices-to-job-candidates-new-york-city-aims-to-regulate-the-use-of-artificial-intelligence-in-the-workplace/>.
- O’Neil Risk Consulting and Algorithmic Auditing (ORCAA). (2020). *ORCAA’s Algorithmic Audit of HireVue—Description of Algorithmic Audit: Pre-built Assessments*. <https://www.hirevue.com/resources/orcaa-report>.
- Orwat, C. (2020). *Risks of Discrimination through the Use of Algorithms* (p. 122). Federal Anti-Discrimination Agency (FADA). www.antidiskriminierungsstelle.de.
- Raab, C. (2020). Information privacy, impact assessment, and the place of ethics. *Computer Law & Security Review*, 37, 105404. <https://doi.org/10.1016/j.clsr.2020.105404>
- Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, <https://doi.org/10.1145/3351095.3372828>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, <https://doi.org/10.1145/3351095.3372873>
- Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018). *Artificial Intelligence & Human Rights: Opportunities & Risks* (SSRN Scholarly Paper ID 3259344; p. 63).
- Harvard University, Berkman Klein Center for Internet & Society. <https://doi.org/10.2139/ssrn.3259344>
- Raub, M. (2018). Bots, bias and big data: artificial intelligence, algorithmic bias and disparate impact liability in hiring practices comment. *Arkansas Law Review*, 71(2), 529–570.
- Research Centre of the Slovenian Academy of Sciences & Arts. (2017). Satori Policy Brief: Supporting ethics assessment in research and innovation (p. 8). European Commission. https://satoriproject.eu/media/SATORI-policy-brief-_2017_Supporting-ethics-assessment-_26-06-2017.pdf.
- Robertson, K., Khoo, C., & Song, Y. (2020). To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada. Citizen Lab and International Human Rights Program, University of Toronto. <https://citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf>.
- Schellmann, H. (11 February 2021). Auditors are testing hiring algorithms for bias, but there’s no easy fix. *MIT Technology Review*. <https://www.technologyreview.com/2021/02/11/1017955/auditors-testing-ai-hiring-algorithms-bias-big-questions-remain/>.
- Scherer, M. (2017). AI in HR: Civil rights implications of employers’ use of artificial intelligence and big data. *Scitech Lawyer*, 13(2), 12–15.
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: challenges and a path forward. *California Management Review*, 61(4), 15–42. <https://doi.org/10.1177/0008125619867910>
- The Leadership Conference Education Fund. (2020). Civil Rights Principles for Hiring Assessment Technologies (p. 6). <https://civilrights.org/resource/civil-rights-principles-for-hiring-assessment-technologies/>.
- United Nations. (2012). The Corporate Responsibility To Respect Human Rights - An Interpretive Guide. https://www.ohchr.org/Documents/Publications/HR.PUB.12.2_En.pdf.
- United Nations Human Rights Regional Office for Europe. (2018). *Make A Difference: An Introduction to Human Rights* (p. 205). United Nations.
- Venkatasubramanian, S., & Alfano, M. (2020, January). The philosophical basis of algorithmic recourse. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 284–293). https://europe.ohchr.org/Documents/Publications/MakeADifference_EN.pdf.
- Wright, D., & Friedewald, M. (2013). Integrating privacy and ethical impact assessments. *Science and Public Policy*, 40(6), 755–766. <https://doi.org/10.1093/scipol/sct083>
- Yeung, K. (2018). *A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework* (SSRN Scholarly Paper ID 3286027). Social Science Research Network. <https://papers.ssrn.com/abstract=3286027>.
- Zuloaga, L. (11 January 2021). Industry Leadership: New Audit Results and Decision on Visual Analysis. *HireVue*. <https://www.hirevue.com/blog/hiring/industry-leadership-new-audit-results-and-decision-on-visual-analysis>.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.