Artificial Intelligence in hiring: friend or foe?

ai:conomics
policy*brief*

May 2022

Artificial Intelligence in Hiring: Friend or Foe?

Fleck, L., Rounding, N., and Özgül, P

Key messages

- The inclusion of artificial intelligence (AI) in hiring practices can have important implications on the extent and magnitude of discrimination in labour markets. Persistent bias can further contribute to rising inequalities, skill mismatch and inefficient labour market outcomes.
- While AI can potentially reduce discrimination by increasing the diversity of successful candidates due to its ability to predict and reach objective decision-making, this does not hold for each case in practice.
- Under certain circumstances, AI can further exacerbate worker discrimination by replicating human biases and producing unfair outcomes at the expense of certain individual groups.
- Outcomes of AI-based hiring highly depend on the specific algorithm design. If a poor algorithm design can manifest in poor hiring decisions, it is a crucial aspect to factor in when introducing AI to the hiring process.
- Workers' perception on the fairness of AI might have a stronger influence on the hiring process and self-selection of workers than the actual fairness of AI. Worker perception therefore becomes a critical ingredient when making informed choices about algorithm designs as well as policies for its ethical implementation.

1. Introduction

Being a powerful predictive technology, artificial intelligence (AI) can be adopted by firms to increase efficiency within various processes, one of them being the hiring practices in human resource management (HRM) (Bogen & Rieke, 2018; Gonzalez et al., 2019; Tambe et al., 2019). As there are a multitude of different opinions and concerns about AI implementation[1] in hiring, we aim to focus on one specific debate: whether it has the potential to increase or reduce discrimination. Discrimination within hiring is a long-studied phenomenon, that is often based on recruiters' unconscious, interpersonal bias, and subjective judgements about people based on different personal characteristics[2]. In a larger scheme, discrimination contributes to inequalities in the labour market (Banks & Ford, 2008) by granting access to attractive employment opportunities for individuals from overrepresented groups while simultaneously restricting the access for other groups. Discrimination also leads to skills mismatch and inefficient allocation of resources (McGowan & Andrews, 2015). AI technology does not naturally have preferences for hiring a candidate with a certain appearance, a specific name hinting at a

1. To find out more on the current debate on AI implementation and its overall effects on labour markets, see the [first ai:conomics policy brief](#).

2. See Box below for more in-depth discussion on discrimination in labour markets.

nationality, or a degree from a certain college, among others. One of the strongest potential advantages of AI technology compared to humans is that, under some pre-conditions, it can produce seemingly objective and reliable decisions. Algorithms are able to model predictions based on masses of data that humans do not have the capacity to comprehend in its entirety (Agrawal et al., 2019; van den Broek et al., 2021). However, aspects that remain questionable when considering introducing AI to the hiring process, are not only whether these claims are actually supported by evidence but also how affected people perceive algorithmic decision-making (Gonzalez et al., 2022).

In this policy brief we outline the scientific evidence on the interplay between AI and hiring outcomes such as the quality and diversity of hires. Next to providing some evidence of AI's ability to reduce discrimination and allow for favourable hiring outcomes in Section 2.1, we use Section 2.2 to shed light on a pressing issue in the discussion of AI in hiring: AI bias. In Section 3 we highlight the importance of workers' perceptions of algorithmic decision making and demonstrate the potential consequences this might have for individual- and firm-level labour market outcomes. Lastly, in Section 4 we discuss our findings' implications for regulatory frameworks and policies to ensure fairness throughout algorithmic hiring processes.

What is discrimination?

Discrimination is the unfair or prejudicial treatment of people and groups based on characteristics such as race, gender, age, or sexual orientation (APA, n.d.). The unfair treatment is based on the fact that two individuals are treated differently based upon characteristics that do not contribute to their productivity (Becker, 2010). To demonstrate this concretely, a prejudiced recruiter may choose a white candidate over a black candidate, despite the latter having a better resume or interviewing skills. Researchers have identified discriminatory behaviour in the hiring process, finding that white candidates were more likely to receive interviews and job offers than black candidates with similar characteristics (Bendick et al., 1994). The bias can already occur during CV-screening and some studies show that switching names that hint toward a certain nationality or race could increase interview call back rates significantly (Bertrand & Mullainathan, 2004; Kaas & Manger, 2012; McGinnity et al., 2009). The existence of such bias within the hiring process can have a significant impact upon individuals' earnings, their career trajectory, and the development of their human capital. Additionally, it is an inefficient market outcome due to an inefficient allocation of skills which could, potentially hamper economic growth (McGowan & Andrews, 2015). These considerations make it crucial to identify how and to what extent discrimination exists, at what stage it appears and how it can be mitigated.

2. Can AI reduce discrimination?

Research has repeatedly discovered that hiring is often subject to discrimination based on candidates' personal characteristics that are actually unrelated to their productivity; race or nationality, derived from names or appearance, could be such characteristics (Becker, 2010; Bendick et al., 1994; Bertrand & Mullainathan, 2004; Kaas & Manger, 2012). Recently, researchers have been suggesting the integration of AI into the hiring process to overcome discrimination caused by unconscious biases of humans (van den Broek et al., 2021). Considering that humans only have a comparably limited capacity to process information from different sources, AI can provide more objective knowledge: intelligent hiring algorithms draw on large datasets from diverse data sources such as CVs, interviews, and social media, enabling them to predict which candidate will fit the respective vacancy the best (Agrawal et al., 2019), thereby often exceeding domain experts' insights (van den Broek et al., 2021).

The suitability of AI to the hiring process is reflected in the recent surge in suppliers offering AI solutions for hiring (Raghavan et al., 2020). AI-based systems can have different functions at different stages of the hiring process, e.g., removing biased wording from job advertisements, screening CVs, analysing video-interviews, or evaluating the person-job fit (Gonzalez et al., 2019; Raghavan et al., 2020). The numerous possibilities of applying AI to the early stages of the hiring process emphasize that algorithms have the potential to substantially impact what types of candidates end up in the overall pipeline (Bogen & Rieke, 2018).

Currently, the prime organisational reasoning to implement algorithmic hiring is to increase efficiency and optimize recruitment metrics such as: time to hire, cost per hire, quality of hire, and employee retention (Bogen & Rieke, 2018; Gonzalez et al., 2019). Furthermore, AI's ability to reduce interpersonal bias (van den Broek et al., 2021) contributes to organisational objectives related to promoting diversity and inclusion (Bogen & Rieke, 2018). Practitioners' journals such as Forbes (Windley, 2021), Harvard Business Review (Chamorro-Premuzic & Akhtar, 2019) and Business Insider (Garfield, 2017) have acknowledged the benefits of using AI in hiring while promoting a cautious approach when it comes to relying on algorithms too much, implying that there are two sides to the story. Accordingly, academia has begun to investigate whether AI actually reduces discrimination in hiring or potentially even exacerbates it. The following sections elaborate on identified opportunities and threats of using AI in hiring.

2.1 AI vs. humans: who achieves better hiring outcomes?

Within the hiring process, CV screening is an area that has been identified as a source of discrimination (e.g., Bertrand & Mullnaithan, 2004). Two studies (Cowgill, 2020; Li et al., 2020) have designed AI algorithms to screen CVs and compared them to human performance. Li et al. (2020) view hiring as a bandit problem where a decision maker must balance exploration with exploitation. In this context hiring firms attempt to balance selecting candidates from groups with proven track records (exploitation) against sourcing from underrepresented groups (exploration). Current hiring algorithms based on training data predict positive outcomes for those groups with a proven track record, disregarding candidates from groups that might have untraditional backgrounds but may still excel in the job. Including this exploration into hiring algorithms may be one step toward a more inclusive approach (Li et al., 2020) which is in line with the notion of building a more diverse pool of candidates (Bogen & Rieke, 2018).

Accordingly, Li et al. (2020) compare human recruiters with three algorithms: two trained by supervised learning and one incorporating an increased degree of uncertainty, i.e., an "exploration bonus". The findings show that the exploration algorithm results in substantially higher shares of underrepresented groups in the selection process. Not only does this algorithm perform significantly better than the human, but it also performs better than the two supervised algorithms. An important result to note is that both the supervised learning algorithms designed in this study performed weaker than humans when it came to select a diverse group of candidates, highlighting the potential of algorithms to further entrench biases, discussed below in more detail. The authors highlight the importance of algorithm design to its performance: not all algorithms will perform the same, and this is an important factor to consider when faced with a market of choices.

In an experimental hiring setting, Cowgill (2020) finds that an algorithm selected CVs of candidates who were better performing, more likely to pass interviews and more likely to receive job offers. Furthermore, the algorithm increased the number of non-traditional hires, including women, racial minorities, candidates without referrals, from non-elite colleges, and those with no prior work experience. Superficially, it seems the algorithm was trained with similar methods (i.e., LASSO and other standard machine learning techniques) to the supervised algorithms in Li et al. (2020), who found that these algorithms could further entrench bias. The contrasting results indicate a need for future research to build a consensus on the performance of algorithms. Incorporating different algorithmic designs, such as the exploration algorithm used by Li et al. (2020), could lead to more in-depth analyses and provide more insight into how discrimination could be reduced. Furthermore, as highlighted by Cowgill (2020), the composition of the training dataset could lead to differential results. In his models, the author demonstrates how more or less statistical noise in decision making could impact the results. Therefore, it is of vital importance to replicate the results of such studies using differently composed training data.

Beyond CV screening, AI could also be used to analyse job suitability tests. These are questionnaires sent out prior to selection for recruiters to factor into their decisions. Hoffman et al. (2018) test the outcomes of such an AI and find positive effects on the tenure of successful hires vis-a-vis the previous human-led regime. Moreover, they investigate if the quality of hires is higher when recruiters follow the algorithm's recommendation or when they actively overrule its recommendations. They show that recruiters that overrule the algorithm make worse decisions, their selected workers have shorter employment durations and were not more productive. Their study contributes to the evidence that an AI may offer an improvement over a human recruiter: not only at identifying better performing candidates but identifying a more diverse set of candidates as well. The employed algorithm was proprietary (Hoffman et al., 2018) and there is little discussion of its design, or what method of learning the algorithm uses. Revealing the model technology and the data utilized to train the model would be necessary for a final, objective assessment of the AI technology, as discussed above.

2.2 Caution: Not every AI is unbiased

In the above section we have outlined some cases where an AI could improve upon a human decision maker and reduce discriminatory outcomes in the hiring process. However, we have also demonstrated that different algorithms can result in different outcomes. As highlighted by the supervised learning algorithms in Li et al. (2020), there is a risk that an algorithm could further entrench bias. Combined with this, there exist multiple media stories highlighting the potential for AI algorithms reinforcing bias instead of reducing it as well as studies highlighting AI bias in other areas, such as facial recognition tools (Leslie, 2020). Therefore, this section is devoted to uncovering the underlying sources of AI bias, together with examples highlighting how AI bias emerges.

Current AI algorithms rely heavily on training data created by humans, and predominantly learn using a combination of several statistical approaches applied to the training data. This method of learning is generally labelled machine learning or deep learning (OECD, 2019). A recent literature study (Cecere et al., 2021) identified the root causes of AI bias as mainly arising from the use of (i) *unrepresentative data* and *training samples*, (ii) *mislabelled outcomes* and (iii) *biased programmers* (Cecere et al., 2021; Cowgill, 2020; Cowgill & Stevenson, 2020; Cowgill & Tucker, 2020). Algorithms are not created intelligent; they have to be trained on data before being applied to the real world. The algorithm takes this data and assumes that it is representative of the real world, and hence assumes that its decisions are applicable to both the training data and the real world. If the training data is biased or unrepresentative, the algorithm could naively replicate the selection processes that created the unrepresentative data, therefore entrenching the bias (Cowgill & Tucker, 2020; Kleinberg et al., 2018). Moreover, since AI algorithms tend to be trained, and later learn from individuals' past behaviours and characteristics, discrimination is likely to occur in a vicious cycle (Cecere et al., 2021). In the context of hiring, if the AI training data had been built on historical employment data, which, e.g., contains an implicit bias favouring white men over Afro-American and Hispanic employees, then the algorithm may detect such patterns, and disregard certain groups of applicants. The decision-making could then be in favour of the overrepresented group, who is historically more likely to be chosen for a job interview (Köchling & Wehner, 2020; Tambe et al., 2019). Global tech giant Amazon serves as a prime example in this scenario: In 2018, the company discovered that its AI-powered recruitment system was based on historical job performance data, which was severely male-dominated, and in return, contained higher performance scores for white men. As a result, trained on that selection of information, the algorithm gave higher scores to white male applicants, while it was selecting out women and candidates with attributes associated with women (Tambe et al., 2019).

Algorithmic distortion can also be linked to mislabelling of some outcomes in the training data. This might arise due to subjective tastes of human decision-makers or failed statistical reasoning (Cowgill & Tucker, 2020). In a recent study, Obermeyer et al. (2019) highlight the importance in choosing relevant labels to train the algorithm by documenting important algorithmic disparities that take place in the healthcare sector. They show that at the same level of algorithm-predicted risk, indicating the healthcare needs, less healthy black patients received similar risk scores to healthier white patients, while in fact, they exhibit more severe cases of diabetes, high blood pressure, cholesterol, and anaemia. The reason for such racial bias is because the healthcare algorithm's prediction is based on healthcare costs rather than illnesses, and since less money is spent on black patients who have the same level of need, the algorithm falsely concludes that black patients are as healthy as the white patients (Obermeyer et al., 2019). This example shows a potential consequence that might arise as a result of incorrectly labelled outcomes; the training data labels low-cost patients as healthy, while in some cases these low costs are due to unequal access to healthcare and socio-economic status (Cecere et al., 2021; Cowgill & Tucker, 2020; Obermeyer et al., 2019).

AI developers and programmers who oversee the writing, labelling, and training of the algorithms might also have an influence in generating algorithmic biases, especially if they lack technical knowledge on AI ethics or have significant biases themselves. The existence of such "biased programmers" may produce unrepresentative training samples, mislabelled outcomes, and/or they might fail to pay adequate attention to available training examples and measures of accuracy during the development process (Cowgill & Tucker, 2020).

Attempting to investigate the formation and prevalence of main sources to AI bias, Cowgill et al. (2020) study whether AI bias more likely results from biased training data or biased programmers. The results of their field experiment demonstrate that biased training data is the main contributor to biased predictions, and that AI developers are highly representative, and might exhibit biases that are transferred to the algorithms they develop (Cecere et al., 2021; Cowgill et al., 2020). In short, while each source of AI bias is case-specific, the underlying mechanism stems from the fact that human biases carry over to AI-based approaches. AI algorithms produce discriminatory and biased outcomes if their trained data and samples are either inaccurate or biased (Köchling & Wehner, 2020; Mujtaba & Mahapatra, 2019). If the data consists of some degree of cultural or personal bias, then an AI algorithm can inherit and replicate the bias, producing unfair outcomes, such as under- or overrepresentation of certain population groups. The resulting discriminatory outcome therefore raises questions on the trustworthiness of AI systems (Lee, 2018; OECD, 2019).

In light of the above discussions, understanding the training data fed to the AI carries great importance to gauge its ability to reduce or exacerbate discrimination. If an AI is trained on discriminatory data, the above evidence points to the fact that it would further entrench biases. This suggests that alongside the algorithm design choices, the existing data is critical to assess when choosing an AI tool, as the different levels of bias within the training data necessitate different algorithmic approaches.

3. Perceptions of algorithmic decisions: trust vs. doubt?

While the evidence laid out above reviews AI as a tool to reduce discrimination, the use of AI in the recruitment process raises other concerns beyond discrimination. Although AI can reduce discrimination and this reduction can be accurately tested, it may still be an inappropriate tool for hiring. One major concern is whether individuals who are subjected to algorithmic decisions perceive them to be fairer than human decisions, irrespective of whether they are actually fairer or not (Lee, 2018). This human perception is important for two reasons, firstly if a candidate perceives an algorithm, and hence the recruitment process, to be unfair, then they may not apply or accept an offer for a position. Such self-selection of workers can have a negative impact on the hiring organisation as the candidate may have led to greater productivity gains than an alternative, leading to lower overall output and growth. The alternative candidate may be so ineffective that they need to be replaced, further driving up costs through increased employee turnover and inefficient skill allocation. Secondly, if a candidate decides to join an organisation despite having misgivings about the hiring process, this could lead to a breakdown in trust between the employer and the employee. This lack of trust can have an impact on job tenure and consequently cause employee turnover, which is considerably costly for organisations, estimated upwards of $16 billion per year for American companies (Scott, 2017). Therefore, ensuring candidates' satisfaction with the recruitment process and trust in the fairness of the hiring decision can be seen as a net benefit to organisations.

Given this, it is pertinent to discuss whether individuals' perceptions of hiring algorithms are positive or negative, and the potential mechanisms behind these perceptions. Newman et al. (2020) find that hiring decisions made by an AI are perceived to be less fair than identical human decisions. Similarly, Acikgoz et al. (2020) show that interviews conducted by an AI are perceived to be less fair than ones conducted by a human. Adding to this, Gonzalez et al. (2019) find that workers have less trust in organisations that use algorithms as decision makers in hiring. Gonzalez et al. (2022) outline potential mechanisms for the trust gap between human and algorithmic decision making. Applicants often have limited information on how an AI is used in a particular decision-making process or may not be able comprehend or process such information if it is available. They will then use simple mental heuristics to evaluate the relative fairness, potentially drawing upon media stories which may promote rather negative outcomes. Applicants may also perceive AI decisions to be less fair due to an unfamiliarity with AI itself and the decision-making process in which the AI is embedded, whilst they may have a better understanding of a human based decision process, thus providing them with more information to evaluate the fairness.

These potential mechanisms may be alleviated with greater transparency about the use of AI. Whilst a firm may easily be able to explain where in a process an AI is implemented and whether it has the final decision power, it is more difficult for candidates to comprehend on what grounds an AI comes to a decision. Modern AIs are trained using neural networks and deep learning techniques, these are opaque techniques as it is not known exactly how an AI comes to its decisions. Researchers are working towards explainable AI (see Arrieta et al., 2020 for a discussion) to allow for a deeper understanding of its respective decision processes. This is important to develop if AI is to become 'trustworthy'.

4. Policy implications

What are the current regulatory efforts to ensure AI is human-centric, trustworthy, and safe? In its proposed AI Act (European Commission, 2021) the European Union classifies the use of AI for recruitment as high risk. Under this classification, an AI technology used for recruitment is not banned but rather is regulated and has to fulfil a number of conditions such as creating a risk management system, transparency conditions and human oversight. The risk management system requires the identification and analysis of the known and foreseeable risks, estimation and evaluation of these risks, and possible risks identified in post-market monitoring. The possibility of AI naively replicating human bias is one of the primary risks raised in this policy brief and is one that would have to be considered under this legislation. Tackling another one of the concerns raised by this policy brief are the transparency requirements, requiring that users be able to interpret the system's output and comprehend its intended purpose. This works towards allaying some of the concerns regarding people's perceptions of AI and its relative trustworthiness. If these transparency requirements are met, it may contribute towards building trust, however, this is not guaranteed and future work identifying whether these requirements do so and what else can be done to build trust is necessary.

There are reasons to believe that the AI Act works towards some of the challenges outlined above. One of the primary questions raised by this legislation is whether the costs stifle the innovation and use of AI algorithms through an overly heavy administrative burden (Czarnocki, 2021). In the other direction, there are concerns that policies such as the AI Act requiring some forms of auditing do not go far enough to weed out harmful tech. Sloane (2022) highlights this when discussing a New York City legislation requiring auditing of hiring algorithms, arguing that these audits do not sufficiently assess the suitability of tech. These are debates that will continue elsewhere but need to be considered when assessing the impact of policies on AI in hiring.

Given that the AI Act has not gone into legislation yet, it is too early to adequately assess its impact, however, it is an interesting and important development that will heavily impact the future trajectory of AI in the European Union.

5. Conclusion

The above discussions highlight that AI has great potential as a tool to reduce discrimination in hiring. They provide suggestive evidence that an AI can increase the diversity of successful candidates whilst also leading to better performing candidates overall. However, well documented use cases, such as the above-mentioned discriminatory recruiting algorithm at Amazon, highlight the need to act cautiously. As described, the way that AI currently learns, has a strong element of human interaction. Whether this be in curating the training data used by the AI, labelling this data or labelling outcomes, bias can transfer from the human to the AI. The AI can consequently further entrench biases, maintaining or even exacerbating discrimination. The results of the supervised algorithms in Li et al. (2020) hint at this possibility, with the better performing exploration algorithm designed specifically to select a wider range of candidates. This highlights the importance of design choices to the performance of algorithms: if designed poorly, an AI could naively replicate unconscious biases displayed by humans, thus defeating its purpose of objective hiring.

As shown by the research into individuals' perceptions of AI in hiring, individuals may not trust decisions made by an AI, even if those decisions are as fair or fairer than those made by a human. This result highlights the need for transparency wherever an AI is used. Humans tend to mistrust the unknown, which leads us to the conclusion that to achieve greater acceptance toward and participation in algorithmic hiring, transparency and openness are crucial. Whilst the explainability of AI is still in its infancy, transparency from users and vendors of AI recruitment tools about the design choices and the training data used would help to build trust between humans and machines. Technologies do not develop deterministically; AI is not going to instinctively evolve to become more or less discriminatory. Within each technological trajectory there are numerous choices that need to be made about the design and use of an artifact. Therefore, AI technologies will be more transparent and, thus, more explainable by cautiously made decisions that are explicitly documented in detail and made available for the people concerned.

This policy brief has attempted to shed light on the role of AI in hiring and some of the crucial choices that can be made with this technology. When designing or using an AI, it is vital to consider the data that it is trained on, the method used to train the AI, and the selection criteria used.

References

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, *28*(4), 399–416. https://doi.org/10.1111/ijsa.12306

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives*, *33*(2), 31–50. https://doi.org/10.1257/jep.33.2.31

American Psychological Association (APA). (n.d.). *Discrimination: What it is, and how to cope*. Retrieved May 2, 2022, from https://www.apa.org/topics/racism-bias-discrimination/types-stress

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. *Information Fusion*, *58*, 82–115.

Banks, R. R., & Ford, R. T. (2008). (How) does unconscious bias matter: Law, politics, and racial inequality. *Emory LJ*, *58*, 1053.

Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press.

Bendick Jr, M., Jackson, C. W., & Reinoso, V. A. (1994). Measuring employment discrimination through controlled experiments. *The Review of Black Political Economy*, *23*(1), 25–48. https://doi.org/10.1007%2FBF02895739

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*(4), 991–1013. https://www.doi.org/10.1257/0002828042002561

Bogen, M., & Rieke, A. (2018). *Help wanted: An examination of hiring algorithms, equity, and bias.* Upturn. Retrieved April 2, 2022, from https://www.upturn.org/work/help-wanted/

Cecere, G., Corrocher, N., & Jean, C. (2021). *Fair or Unbiased Algorithmic Decision-Making? A Review of the Literature on Digital Economics. A Review of the Literature on Digital Economics. SSRN.* https://ssrn.com/abstract=3943389

Chamorro-Premuzic, T., & Akhtar, R. (2019). *Should Companies Use AI to Assess Job Candidates?* Harvard Business Review. Retrieved April 26, 2022, from https://hbr.org/2019/05/should-companies-use-ai-to-assess-job-candidates

Cowgill, B. (2020). Bias and productivity in humans and algorithms: Theory and evidence from resume screening. *Columbia Business School, Columbia University*, *29*. https://conference.iza.org/conference_files/MacroEcon_2017/cowgill_b8981.pdf

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020). Biased programmers? Or biased data? A field experiment in operationalizing ai ethics. *Proceedings of the 21st ACM Conference on Economics and Computation*, 679–681. https://doi.org/10.1145/3391403.3399545

Cowgill, B., & Stevenson, M. T. (2020). Algorithmic social engineering. *AEA Papers and Proceedings*, *110*, 96–100. https://www.doi.org/10.1257/pandp.20201037

Cowgill, B., & Tucker, C. E. (2020). Algorithmic fairness and economics. *Columbia Business School Research Paper*. SSRN. https://dx.doi.org/10.2139/ssrn.3361280

Czarnocki, J. (2021). *Good intentions, unintended consequences? How the Proposal for the AI Act might kick innovation out of the EU*. CITIP Blog. Retrieved May 3rd, 2022, from https://www.law.kuleuven.be/citip/blog/good-intentions-unintended-consequences/

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council. Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. https://artificialintelligenceact.eu/the-act/

Garfield, L. (2017). *A startup claims to have finally figured out how to get rid of bias in hiring with artificial intelligence*. Business Insider. Retrieved April 28th, 2022, from https://www.businessinsider.com/hiring-diversity-brain-games-artificial-intelligence-automation-2017-9

Gonzalez, M., Capman, J., Oswald, F., Theys, E., & Tomczak, D. (2019). "Where's the I-O?" Artificial Intelligence and Machine Learning in Talent Management Systems. *Personnel Assessment and Decisions*, *5*(3). https://doi.org/10.25035/pad.2019.03.005

Gonzalez, M. F., Liu, W., Shirase, L., Tomczak, D. L., Lobbe, C. E., Justenhoven, R., & Martin, N. R. (2022). Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. *Computers in Human Behavior*, *130*. https://doi.org/10.1016/j.chb.2022.107179

Hoffman, M., Kahn, L. B., & Li, D. (2018). Discretion in hiring. *The Quarterly Journal of Economics*, *133*(2), 765–800. https://doi.org/10.1093/qje/qjx042

Kaas, L., & Manger, C. (2012). Ethnic discrimination in Germany's labour market: A field experiment. *German Economic Review*, *13*(1), 1–20. https://doi.org/10.1111/j.1468-0475.2011.00538.x

Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic fairness. *AEA Papers and Proceedings*, *108*, 22–27. https://www.doi.org/10.1257/pandp.20181018

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, *13*(3), 795–848.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1). https://doi.org/10.1177%2F2053951718756684

Leslie, D. (2020). *Understanding Bias in Facial Recognition Technologies: An Explainer.* The Alan Turing Institute. https://doi.org/10.5281/zenodo.4050457

Li, D., Raymond, L. R., & Bergman, P. (2020). *Hiring as exploration*. (NBER Working Paper Series, No. w27736) National Bureau of Economic Research. https://doi.org/10.3386/w27736

McGinnity, F., Nelson, J., Lunn, P., & Quinn, E. (2009). *Discrimination in Recruitment: Evidence from a field experiment*. (Equality Research Series) Equality Authority; The Economic and Social Research Institute. Retrieved May 15, 2022, from http://hdl.handle.net/10147/76641

McGowan, M. A., & Andrews, D. (2015). *Labour market mismatch and labour productivity: Evidence from PIAAC data*. OECD. https://doi.org/10.1787/5js1pzx1r2kb-en

Mujtaba, D. F., & Mahapatra, N. R. (2019). Ethical considerations in ai-based recruitment. *2019 IEEE International Symposium on Technology in Society (ISTAS) Proceedings*. https://www.doi.org/10.1109/istas48451.2019.8937920

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453.

OECD. (2019). *Artificial Intelligence in Society*. https://doi.org/10.1787/eedfee77-en

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*, 469–481. https://doi.org/10.1145/3351095.3372828

Scott, A. (2017). *Tech Leavers Study: A First-of-its-kind Analysis of why People Voluntarily Left Jobs in Tech*. Kapor Center for Social Impact.

Sloane, M. (2022). To make AI fair, here's what we must learn to do. *Nature*, *605*(7908). https://doi.org/10.1038/d41586-022-01202-3

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *California Management Review*, 28. https://doi.org/10.1177%2F0008125619867910

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly*, *45*(3), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

Windley, D. (2021). *Is AI The Answer To Recruiting Effectiveness?* Forbes. Retrieved April 26, 2022, from https://www.forbes.com/sites/forbeshumanresourcescouncil/2021/06/16/is-ai-the-answer-to-recruiting-effectiveness/?sh=431b6ff2d7c7