Algorithmic evaluation in the recruitment process – does it increase diversity in organizations?

A work in progress[1]

Moa Bursell
Institute for Futures Studies

Lambros Roumbanis
Stockholm Centre for Organizational Research

Introduction

Most experts agree that a continued implementation of artificial intelligence (AI)[2] into different spheres of social life will have a dramatic impact on decision-making across a wide range of organizational contexts. Public as well as private sector actors are already in the process of transforming their decision-making procedures through the support from algorithms that collects and bridges data, evaluates, and creates systematic recommendations (Bader and Kaiser 2019; Beer 2017; Burrell and Fourcade 2021; Jarrahi et al. 2021; Kiviat 2019; von Krogh 2018; Zarsky 2016). However, all technological changes can have both intended and unintended consequences, and it is very difficult to foresee how such changes will interact with pre-existing social institutions and societal distributions of power and resources. Thus, while there is expert consensus that AI systems and algorithmic evaluations already are

[1] This is the first draft of this study, and the results are therefore preliminary.

[2] We will use the term AI to cover all sorts of complex systems and smart tools that are invented by humans to solve specific tasks that only humans otherwise would have been able to perform, like playing a game of chess, screen applicants for an insurance, or calculate astronomical phenomena. There is a common misunderstanding among many laypersons to think of AI as something that necessarily must resemble human intelligence, whereas the term artificial intelligence is used to denote a variety of different types of intelligences that are *not* natural like the intelligence of an octopus or a bat, but artificial like the computer program AlphaZero. The term often used when researchers discuss the possibility to create intelligent machines with the human-like capacity to solve many different tasks, is “artificial general intelligence” (see e.g., Fazi 2021; Russell 2019). Whether it is plausible or not to assume that such a system will ever be able to have subjective beliefs, deep intuitions, and existential anxiety like human beings, is something that philosophers and computer scientists have discussed since at least Alan Turing’s days (1950; see also McCarthy 1979; Searle 1980; Sloman 2014; Larson 2021).

beginning to transform both private and public organizations, there is a rather wide disagreement on the nature and consequences of these changes.

Several scholars emphasize the various ways in which these emerging technologies will improve societies and facilitate everyday life for its citizens (Jaton 2021; Kahneman et al. 2021; Kleinberg et al. 2019; Logg et al. 2019; Makridakis 2017; Sunstein 2019). The most common argument among this group of scholars is that well-designed and sensitively coded algorithms will simply outperform human judgment in being more objective, efficient, and unbiased. Even highly skilled experts often disagree on how to define, for example, scientific quality and the future potential in new research proposals; panel group decisions are influenced by seemingly unavoidable measures of arbitrariness and chance (Roumbanis 2017, 2022). With the implementation of algorithms in the evaluation process, human bias and prejudice, can be systematically reduced. However, other scholars identify problems and risks with letting over some of the control to algorithms, such as enhanced inequality and discrimination of already disadvantaged groups (Ajunwa 2020; Diakopoulos 2016; Howard and Borenstein 2018; Köchling and Wehner 2020; Obermeyer et al. 2019; Zarsky 2016). In the present study, we will focus on the consequences that the implementation of an algorithm-based recruitment process can have for facilitating the inclusion of disadvantaged groups and the promotion of a general diversity on the Swedish labor market.

Recruitment systems based on AI has generally been depicted as a game-changer when it comes to human resources management. The driving force among organizations for making decisions based on algorithms is that it saves time and reduces costs, enhances productivity, and increases certainty and control (Köchling and Wehner 2020; Ragahavan et al. 2019). The use of algorithms is also assumed to improve fairness and impartiality by removing human prejudices and subjective preferences from at least the early phase of the recruitment process (Kahneman et al. 2021; Kleinberg et al. 2019; Raghavan et al. 2019; Sunstein 2019). These kind of recruitment systems can support the expert recruiter by identifying the best job candidates based on their real merits, candidates that they would perhaps have missed if they only relied on their own professional knowledge and “gut feelings” (Moss and Tilly 2001; Rivera 2015). But depending on how the algorithms are designed and how they merge with organizational practices and routines, this is far from the only possible outcome; algorithms can also reproduce and exacerbate existing patterns of exclusion (Ajunwa 2020; Eubanks 2018; O’Neill 2016).

The implementation of AI recruitment systems in the hiring processes has thus far gained rather limited scholarly attention (Vrontis et al. 2021; Upadhyay and Khandewal 2018; van Esch et al. 2019). Due to huge gains in efficiency and productivity for HR departments, AI recruitment is likely to be widely implemented and further developed, starting with larger companies and recruitment agencies, i.e., companies that recruit a lot of people and process a large amount of job applications (Köchling and Wehner 2020). If this development continues, algorithms will replace humans in a significant share of the recruitment process. This technological transition has been set in motion, even though we have a rather limited knowledge about its social consequences. As recently highlighted by Howard and Borenstein (2018: 1521): “These specialized AI algorithms have been liberated from the minds of researchers and startups and released onto the public. Yet intelligent though they may be, these algorithms maintain some of the same biases that permeate society.” Given the potential impact of this development for labor markets already characterized by rather severe social exclusion, inequality, and discrimination (Bursell 2014; Keeley 2015), studying the consequences of different organizations that implement AI systems is of tremendous societal importance.

The academic discussion thus provides both optimism and pessimism regarding the possible impact that algorithms might have in the future. However, there is yet no solid empirical foundation of research to support pushing these interesting discussions in new directions. Perhaps more fruitful question to pose, in this context, are: *when* does replacing humans with algorithms lead to better assessments and decisions*, when* does it have negative effects, and for *whom?* Addressing these questions will assist us in understanding what type of tasks, if any, that should be delegated to algorithms. In this paper, we will study this technological transition by comparing differences in outcomes in terms of diversity when job applicants have been assessed by an AI-driven recruitment process and with a traditional, CV-based recruitment process. We will investigate whether an AI-driven process result in a demographically more diverse set of employees than traditional recruitment by analyzing the employment and recruitment records from one of Sweden's largest food retail companies during the period 2019-2020. The study is outlined as follows. First, (i) we introduce a general understanding of algorithms and how they transform the conditions for organizational decision-making. We also discuss the issue of human bias, and the way recruitment algorithms can help to reduce discrimination on the labor market. Second, (ii) we theorize the notion of algorithmic evaluation and introduce the concept of *meta-algorithmic judgment* as a central part of our analytical framework. Third, (iii) we explicate our empirical case, methods, data collection, and results. And fourth (iv) we discuss our findings and try to identify plausible mechanisms for them in our concluding discussions.

The socio-technical power of algorithms

In the history of human civilizations, we find a number of methods that people have used regularly in order to make important decisions. For example, by consulting oracle's, drawing lots, organizing competitions, or adhering to the judgments of knowledgeable experts, uncertainty regarding the best practice has been reduced and decision-making thereby facilitated. In our computerized age, algorithms have come to play a new and increasingly dominating role in how organizations are managed (Jarrahi et al. 2021; Yeung 2018). To support their decision-making processes, many stake-holders are placing their hopes to new systems based on smart algorithms to make their organizations more competitive, efficient, and legitimate, in today’s complex geopolitical and financial landscapes.

Algorithm as a concept has gained great currency as one of the vehicles through which innumerable socio-political and ethical concerns are projected today (Yu 2021; see also Bareis and Katzenbach 2021; Schiølin 2019). But what exactly does this concept refer to? To some extent, the word algorithm has become rather obscured and even mystified, as if algorithms have a life of their own, apart from humans (Lee et al. 2019; Ziewits 2016). But algorithms are, essentially, logico-semantical constructions that must be coded in a specified program language in a computer system. Algorithms are, according to Russell (2019: 34), designed by using simpler building blocks called subroutines, for example, “a self-driving car might use a route-finding algorithm as a subroutine so that it knows where to go. In this way, software systems of immense complexity are built up, layer by layer.” A self-driving car is a good illustration of problem-specific engineering, where algorithms are designed to solve specific tasks, and as with all kinds of technologies, shaped by assumptions about social needs and economic opportunities (Stilgoe 2018). The algorithms used in HR recruitment systems are also designed to solve a specific task, namely, to evaluate and rank the most suitable candidates for a certain job position. When assessing the consequences of the implementation of such algorithms in society, it is also important to keep in mind that in all types of contexts were algorithms are involved, there will also be *human-machine interaction effects*.

Thus, what is important to emphasize when considering different algorithms are their practical consequences, as well as the broader organizational context in which they are implemented and used. There are, for example, significant differences in contextual conditions between a health care algorithm used for disease detection and an algorithm used by the police to reduce crime in a poor neighborhood. All algorithms are constructed to perform something specific, no matter their simplicity or complexity. Machine learning (ML) algorithms have revolutionized certain areas in contemporary societies, being more complex and powerful than conventional rule-based (RB) algorithms. However, they are both human inventions and all algorithms need some kind of human input and management to produce relevant evaluations. Their success will depend on a combination of algorithm capacity, human input, the quality of data and the feasibility of applying it to a specific question or context. A recent case serves to illustrate this issue, a case where the legitimacy of a rule-based algorithm was seriously questioned. In the US, at the Stanford Medical Centre, blame was placed on a "distribution algorithm" when controversy erupted on the misallocation of Covid-19 vaccines to different stakeholders since administrators had been prioritized before the medical staff Jarrahi et al. (2021: 10). This example puts the finger on the real crux of the matter, that is: algorithms are often designed to score and rank people, which always has consequences for their futures. Algorithms are constructed as sorting mechanisms; they are designed, for example, to facilitate the evaluation of a large number of job candidates in the recruitment process, as a first screening that creates a short-list that is automatically delivered to an HR expert. Biases can be built-in during the design of the algorithm, or in the algorithmic evaluation as such. But it can also slip into the process during the interpretation and handling of algorithmic outputs (what we will call *meta-algorithmic judgment*) and influence the final parts of the recruitment process. In our view, the focus on biases must not necessarily reify the general notion of algorithms, as assumed by for example Lee et al. (2020). The way we conceive of algorithmic evaluations, is both as a folding of different valuation practices, but also as a semi-automated process that inevitably reduces and decontextualizes. Our view supports a dynamic understanding of how evaluative processes generate heterogenous outcomes depending on the interplay between social-technical, cultural, and political aspects in concrete organizational settings.

Can algorithms reduce the impact of human biases in the labor market?

Is AI recruitment a solution to the problem of discrimination, or does it lead to a new kind of Pandora's jar where human implicit biases and bad judgments are transformed into algorithmic biases? In the hopeful narrative on AI implementation, algorithms are conceived of as a revolutionizing set of tools that will support human judgment in making much better decisions. If correctly designed and carefully coded, replacing our own natural capacity to think and judge with AI systems may neutralize bias, sustain impartiality, be more efficient and consistent in the processing of large amounts of very complicated tasks, including the assessment of job candidates. Sunstein (2019) highlighted the potential advantages of using algorithms in the American judicial system, "along every dimension that matters, the algorithm does much better than real-world judges" (Sunstein 2019: 500). However, this line of reasoning, that is, how algorithms can reduce biases from the decision-making process and thereby promote fairness, has been questioned from another point of departure.

According to a more dystopian narrative, in the construction of these AI systems and algorithms, there is a significant risk that the implicit biases (e.g., Greenwald and Banaji 1995) that already exists in social life, is transferred into a much more powerful machinery that will make autonomous evaluations partly beyond our control. Thus, algorithmic biases can reproduce and exacerbate already existing patterns of implicit biases and different status hierarchies (e.g., Ridgeway 2019) in social life. This has been of great concern within critical algorithm studies, law, and social psychology, were arguments and empirical examples have shown how AI systems unintendedly can lead to even more systematical discrimination (Ajunwa 2020; Howard and Borenstein 2018; Kiviat 2019), such as the infamous Amazon case, where machine learning autonomously picked up gender bias from the firm's skewed employee gender balance.

Few judgments can be considered completely objective and/or neutral when it comes to recruitment processes. Even the most professional and well-intended HR personnel can have implicit biases and will rely on their expert “gut feelings” (Rivera 2015). Age, gender, sexuality, race/ethnicity, religious believes, class and family background, physical appearance (attractiveness, weight, tattoos etc.) are factors that implicitly can influences recruiting managers or HR staff when deciding whom to employ. It is well known that employment discrimination systematically bars individuals identified with certain social categories, like ethnic or racial minorities, from employment (for a review, see Baert 2018), or channel them to employment within certain types of jobs (Pager, Bonikowski and Western 2009; Bursell, Bygren and Gähler 2021). Thus, employment discrimination creates both labor market exclusion and segregation. Previous research indicates that as much as 90 percent of all discriminatory decisions occur prior to the interview (Bovenkerk 1992). If human biases that only or primarily occur in the earlier phases of the recruitment process can be removed, there is thus significant potential for discrimination to be decreased by automating this part of the recruitment process by delegating it to AI-driven recruitment systems.

How successfully an unbiased algorithm may decrease discrimination and increase diversity hinges, however, at least to some extent on why human recruiters discriminate. When employment discrimination is driven by explicit bias, removing human judgement from the early phases of the recruitment process will likely do little good. In such cases, replacing humans with algorithms will only move the occurrence of discrimination to a later phase of the recruitment process, from the initial screening phase to the interview stage (see e.g., Åslund and Skans 2012). However, employment discrimination is often depicted as a phenomenon that persists despite a hegemonic societal discourse on the undesirability of discrimination. Thus, it may be argued that employment discrimination is rarely rooted in explicit, deliberate recruiter bias. Instead, employment discrimination is believed to be the result of *implicit biases:* associations, stereotypes or beliefs that are activated automatically and non-deliberately, influencing our judgements in everyday life assessments when there is little time for slow and effortful thinking (Bursell and Olsson 2020 2021). These biases may be triggered throughout the recruitment process, but they are believed to be especially crucial in the early screening process – when a large number of job applications are processed and where each application by necessity is assigned very little time, i.e., when our "fast" thinking is at work. The potential impact of implicit bias in this phase of the recruitment process is even more important as firms in the digital age receive an increasing number of job applications, as the low cost of applying online lowers the threshold for applying for jobs.

The concept of meta-algorithmic judgment

In this section, we explore the way algorithms might change the very conditions for organizations to make sensible decisions that have an impact on people's life chances. By introducing the concept of *meta-algorithmic judgment*, our intention is to establish a general term for a sociology of judgment and decision-making that takes into account the innumerable folded relations between algorithmic evaluations and the human judgments. With the prefix *meta*-, which literally means "after," the meaning of the concept is to pinpoint the procedural relationship in time, between algorithmic evaluations and the human judgment that makes the next step in the decision-making process. The overriding question concerns what happens after the algorithms has produced its outcome.

To begin with, most algorithms are constructed to support human judgments in making decisions in different social contexts. Algorithms can both reduce complexities and be more comprehensive; they can also identify hidden patterns in very large amounts of raw data and detect small but important deviations that are unnoticeable to humans. Within health care, for example, algorithms are used to enhance the precision in diagnosing patients and in geoscience, researchers are using algorithms to make more accurate seismological predictions of earthquakes. In these contexts, professionals are making their meta-algorithmic judgments based on very special forms of algorithmic evaluations (which often include Big Data and computer simulations). However, meta-algorithmic judgments are always context-dependent; in fact, there are considerable differences in the scope and limits of interpreting algorithmic outcomes.

In many organizations, algorithms are used because they create opportunities to overcome human biases and other cognitive shortcomings, but also because they are more cost-efficient and powerful in handling data. For the purpose of our study, we will use the concept of meta-algorithmic judgment to emphasize the special conditions of the recruitment process, where the algorithm is scoring and ranking candidates, resulting in a short list. This short list can be seen as the concrete product of the algorithmic evaluation process. Taking this short list as a point of departure, the recruiting manager moves forward making his/her decision about which candidates to invite for an interview. The recruitment algorithm is constructed to screen a large number of job candidates in a more objective, impartial, and efficient way than when HR professionals have to screen hundreds of CVs during a limited period of time. A recruitment algorithm measures how compatible a certain candidate is for a particular job based on personality tests and how well they perform in answering competence-based questions. This marks a standardization of this part of the evaluation process, the motive being to produce a more objective and inclusive short list. Still, one problem with recruitment algorithms seems to be that they are reductionistic. A short list of candidates ranked by the algorithm is not necessarily ideal for the person responsible for making the judgments in the next step in the recruitment process. This is where the issue of meta-algorithmic judgments comes into the picture and becomes a very crucial organizational issue in itself. What happens in this exchange between AI technology and human decision-makers?

In a recent study, Newman et al. (2020) showed how the use of an HR algorithm to evaluate employee performances was perceived by the employees as unfair. Increasing fairness are commonly viewed as key concerns for organizations, yet even though algorithms may remove human biases from sensitive decision-making regarding promotions or gratifications, the process might still be felt as reductionistic by those individuals that are being evaluated. In other words: it can lead the employees to think “that certain qualitative information or contextualization is not being taken into account” (Newman et al. 2020: 149) by the managers. The concepts of meta-algorithmic judgments points to the important step in the evaluation process that follows upon the screening and scoring conducted by the algorithm. Finally, another important issue to consider regarding meta-algorithmic judgments, is the risk of decision-makers of not having sufficient control over the algorithmic evaluation. One problem, for instance, with the lack of control over the selection process, is that it might put recruiting managers in a position where they merely accept algorithmic recommendations without using their own professional judgement (Jarrahi et al. 2021). However, the contrary may also be problematic, that is, if the recruiting manager does not like the recommendations produced by the algorithm, and instead improvises and uses his/her own strategies to choose candidates. In such cases, even if the algorithmic evaluation is unbiased and produces a competence-based shortlist, the outcome of the recruitment process will nevertheless lead to a problematic form of *selection bias*.

The recruitment processes at "FoodMarket"

We analyze employment and hiring records from a large food retail group in Sweden. To keep our case anonymous, we refer to it here as "FoodMarket." The company has an ambitious profile when it comes to diversity goals and equal treatment. It has recently begun the process of automating the initial phases of their recruitment processes through the incorporation of a rule-based algorithm. The goal is to automate all recruitments up until the shortlist stage. From the shortlist, the recruiting manger selects candidates for interview and completes the recruitment process. The implementation of the automated recruitment process began in 2019. In parallel with the automated process, traditional, CV-based recruitment has continued. Thus, the automated process has run parallel to traditional recruitment, enabling a comparison between the two.[3]

The *automated AI process*: Applicants apply for vacancies through FoodMarkets's website. From this website, applicants are redirected to an online platform where they answer questions on qualifying criteria such as previous work experience, education etc. and submits the application. If they are eligible for the position, they automatically receive an invitation to conduct a work test, a test that scores the applicants on how their personality matches the position and their problem-solving capacity. Responses to the basic qualifying questions combined with the outcome of the test results in a score with a maximum of 100. The general principle is that applicants that score below 70 should not be considered.[4]

HR assists the recruiting managers and monitors the process by making sure that applicants receive accurate information on the application process, makes sure that rejected candidates are informed, prepares employment contracts etc. Accordingly, the online job

[3] We have acquired a practical understanding of how the recruitment processes are conducted at the company though meetings and conversations and e-mail correspondence with key informants at the company, and through the reading of formal documents provided by the company.

[4] Reasons for being included even though scoring below 70 include if the applicant is already employed within the organization (i.e., his or her strengths and weaknesses are already known to the employer), or a lack of applicants that score above 70.

application system results in a data set where applicants are terminated at different phases of the recruitment process. Some applicants are rejected automatically at the first stage because they do not meet basic requirements, some in the test phase, some in the shortlist, and some at the interview stage.

In *traditional recruitment,* the job applicants send a personal letter and a CV by email to the recruiting store/work unit. Recruiting managers handle the entire process; she or he screens the applications, creates a shortlist, chooses candidates for interviews, conducts the interviews, and makes the hiring decision. Thus, this process is not centralized and is thus not represented in the centralized online job application system. We refer to this process interchangeably as "traditional" and "CV-based". However, since there is no centralized information about how this recruitment is made, it can also include social network recruitment, and "walk-ins", i.e., cases where applicants show up in person at the store. The central HR department has no data, neither on these processes nor on the job applicants, but they handle employment contracts and registers new employees from these processes in the company records.

Analytical strategy

To study the effect of AI implementation at FoodMarket, we draw on one dataset on employment records of all new employees recruited in 2019-2020, in total 12 336 individuals, and one data set on job applications sent to FoodMarket for the same years. This dataset only includes applications for positions recruited with the algorithmic process. Drawing on the former, we will be able to study potential differences in demographic characteristics between new employees recruited with the two processes. Drawing on the latter, which only includes the algorithmic process, we will be able to identify how different groups fare at different stages of the recruitment process.

The job applications have been stored by FoodMarket in the case of questions or complaints about the employment process, it has never been used by FoodMarket för analysis. We have, in dialogue with FoodMarket about the content of various variables and values, been able to convert this raw data into a coherent, analyzable dataset.[5]

Some job applications have been excluded from the analysis because they have been assigned statuses that are not possible to rank in terms of progress in the application process. These include i) applications that have been withdrawn by the applicant, ii) applicants applying for positions that were withdrawn by FoodMarket, and iii) applicants that have been informed that the position has been filled. Since the system does not record *when* in the process these measures were taken, it is not possible to situate these applications in terms of progress in the recruitment process. In the coding of gender and geographical background, a smaller number of job applications were un-codable due to typing error on behalf of the applicant, such as typing an email address in the field for first name or last name or too many typos/misspellings for name to be intelligible. In total, the dataset consists of 216 929 analyzable observations.

Outcome variables

In the analyses of employment records, the primary outcome variable is *type of recruitment process*: algorithmic or CV-based. In a complementary analysis, we also treat the continuous *tests score* variable as an outcome variable, ranging from 0-100. In the analysis of job applications, the outcome variable is also dichotomous, comparing those who reached at least the interview stage (those who were interviewed, and those who were hired) and those who were not.

[5] A later version of this draft will include an appendix containing more information on the preparation of this dataset.

Explanatory variables

In the analyses of employment records, sex, age, and geographical background constitutes the analyses' explanatory variables. Employees' (juridical) sex is registered in employment records as a dichotomous variable. Age is also registered and is in our analysis coded as above and below 40. The employees' self-ascribed ethnicity or county of birth is not registered in employment records, neither at FoodMarket nor at any other company in Sweden.

We have created a proxy for the employees' *perceived* geographical background based on the employees' surnames using a machine-learning classifier algorithm. For reasons of anonymity and employee integrity, we only had access to employee surname when working with this classification. The job application data, which as mentioned above only includes applications for the algorithmic process, includes, by design, no demographical information on the applicants: the application process should not be influenced by such factors. Thus, the applicants check no boxes for gender, age, and the like. But the applicants' register their first name and their surnames in the applications. Thus, if the application reaches the shortlist, where employers take over the process, most people's sex will be immediately identifiable from their first name, and employers will *perceive* some type of ethnic background from their name, albeit with lower precision than when it comes to gender.

Most recruiters will automatically perceive signals of ethnicity in surnames like "Andersson", "García, or "Mohammed". These perceived backgrounds, whether perceived as ethnicities, cultures, or nationalities, will in turn trigger implicit or explicit attitudes or stereotypes, regardless of whether the applicants identify with this perceived ethnicity or not. Surnames are passed on over generations and may in some cases not reflect the self-ascribed ethnicity of the name-bearer at all. However, self-ascribed ethnicity is not what matters the most in the job application process – it is the signal picked up by the recruiter/employers that influences the chances of being invited to a job interview, not actual ethnicity. In this way, perceived ethnicity may be better predictors of employment discrimination than actual ethnicity, since in the employment process, at least prior to the interview stage, the name is all that the employer see (except in cases where a photograph is attached, but this is not standard procedure in Sweden). Thus, names become a proxy for (perceived) ethnicity.

Lacking data on how surnames are perceived, the specific model used for the classification of geographical background in this study is a neural network consisting of three stacked bidirectional LSTM layers. The classifier model was trained on data from biographical articles of living people on the English-language Wikipedia. The names of the subjects were matched to a country of origin based on the article's belongingness to geographical categories (categories like e.g. "People from Stockholm, Sweden"). Countries around the world were thereafter placed in sub-categories for nine different geographical regions (Scandinavian, Finnish, Eastern European, Spanish, Southern Europe (non-Spanish), Western European, Muslim, African (non-Muslim), and Asian (non-Muslim), with the assumption that the types of names in each of these geographical regions will not overlap to much with those in the other regions. While the dataset is too large to control the quality of the categorizations one by one, we have manually screened through the dataset to ensure that common names are not misclassified and that no significant systematic mistakes have been made.

In the statistical analyses, we have chosen to collapse the nine categories into a dichotomous European-sounding - non-European sounding variable. "European" includes Anglo-Saxon names. "Non-European" includes Spanish names since most people with Spanish names in Sweden have an origin in Latin America (rather than in Spain).

To achieve a proxy for sex, we have built a gender probability estimator based on the Statistics Sweden onomastic database ("SCB namnstatistik") table of first names (*tilltalsnamn*), which contains all given names in Sweden that have at least two bearers and provides numbers for how many women and men, respectively, that have that given name. (This is based on legally registered sex). We have used these data on sex to calculate, for each given name, the probability that the person is a woman using the formula p(woman|name = *x*) = number of women with name *x*/total number of people with name *x*. For more details, see Appendix A. While proxies for sex and geographical background can often be made from names, it is much harder to guess a person's age based on name with any precision. While there are generational trends in naming, these are not distinct enough for the construction of a meaningful proxy for age. We have thus made no such attempt.

Control variables

In the employment records, we control for the geographical location to which the employees have been recruited. We also control for position, through an occupational dummy variable. The employment record contains 203 distinct types of positions, ranging from low- to high skilled, but a vast majority of the positions are low-skilled. As many as 81 percent of the new employees are employed as store employees, and 11 percent as warehouse workers. Thus, the remaining 201 positions are distributed over only 8 percent of the employees. For this reason, we have decided to code the dummy position with separate codes for the two largest categories: store employees and warehouse workers, thereafter categories for: other low-skilled positions, low-skill managers, and high-skilled positions. Test scores, described above, is used as a control variable in the main analyses. In the analyses of job applications, we also control for test score and position and present the models separately for full-time and part-time positions. A control variable for geographic location is not included at this point.[6]

[6] This variable may be included in a later version of the paper, if judged to be of sufficient quality.

Statistical analyses

The analyses of the employment records are conducted with linear probability models (LPM), assessing the probability that new employees are women, have a European-sounding name and are below the age of forty when employed through the algorithmic process compared with the traditional recruitment process. Results are presented with and without fixed effects at the work unit level to control for the grouped nature of the data, i.e., that employments are conducted at different stores/work units across the country. A complementary OLS regression is conducted assessing whether these characteristics predict test scores.

In the analyses of the job application process, we cannot compare the two employment processes as this data only contains the algorithmic process. However, we will be able to identify how far men, women, and job applicants with European and non-European sounding names reach in the algorithmic process. These analyses have been conducted with linear probability models, estimating the probability of being invited to interview.

Results

The newly employed are relatively young, mean age is 27 years old, and 58 percent are women. We code 21 percent as having a non-European sounding family name. 54 percent of the new employees have been recruited through traditional CV-based recruitment; the remaining 46 percent have been recruited through the automated, algorithmic process. Table 1 summarizes descriptive statistics on the new employees. Comparing the two recruitment processes, we do observe mean differences in terms of diversity. In the algorithmic process, there is a somewhat higher share of new employees who are women, who have European sounding surnames, and there is a higher share of younger applicants compared with the employees hired through traditional recruitment.

Table 1. Descriptives by type of recruitment process, new employments

	Algorithmic	Traditional
Number of recruitments	5 653	6 683
Number of locations	177	178
Mean share European names	0.82	0.76
Mean share of women	0.62	0.55
Mean share 40+ years	0.08	0.12
Occupations, column shares		
Store personnel	0.80	0.82
Warehouse workers	0.11	0.11
Other low-skilled	0.04	0.03
Managers	0.04	0.03
High-skilled personnel	0.01	0.01
Year of recruitment, column shares		
2019	0.36	0.64
2020	0.54	0.47

In Table 2, we assess outcomes for algorithmic compared with traditional recruitment in two linear probability models. In model 1, we report probabilities of having a European sounding name, being a woman, and being below the age of 40 in algorithmic compared with CV-based recruitment. In model 2, we include fixed effects at geographical location and control for type of position. The results are stable across the two models and confirm the patterns that were discernible in Table 1. Algorithmic recruitment is associated with a 7-percentage points higher probability of new employees having European sounding names, a 6-percentage point higher probability of being women, and with an 11-percentage points higher probability of being below the age of 40, compared with the employees recruited through CV-based recruitment. Neither controlling for position, nor taking the grouped nature of the data into account, have an impact on the results as evident from model 2. Thus, the occupational or geographical patterns does not correlate in a significant way with recruitment process.

Table 2. LPM models on probabilities of being employed by the algorithmic recruitment process, standard errors in parenthesis

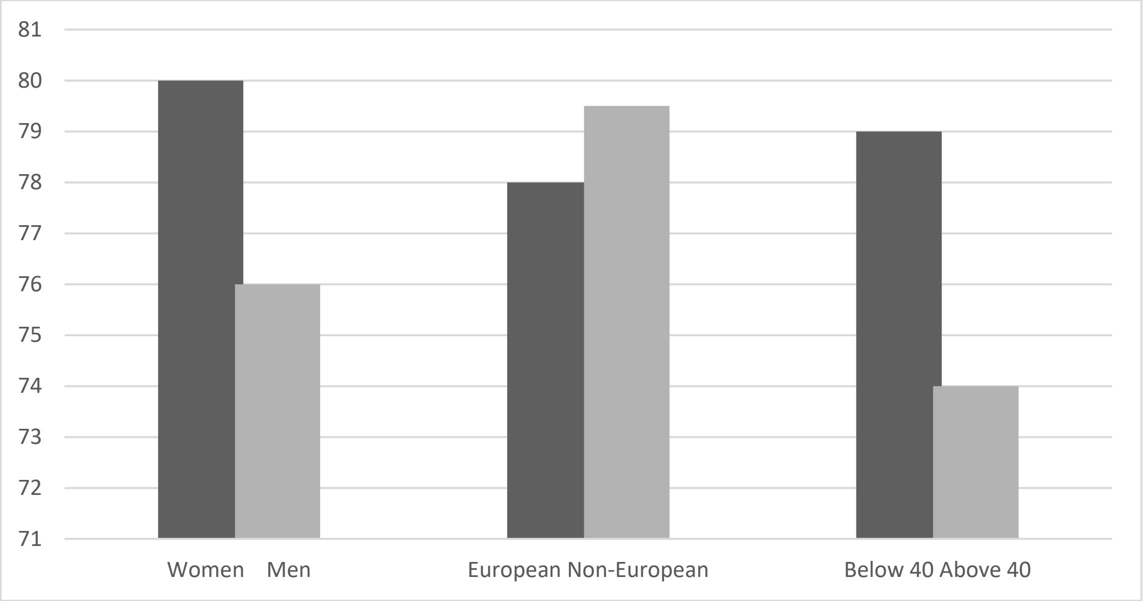
	Model 1	Model 2
European sounding name	0.07***	0.07***
	(0.06)	(0.01)
Female	0.06***	0.06***
	(0.09)	(0.01)
Below forty	0.11***	0.11***
	(0.01)	(0.01)
Position	NO	YES
Constant	0.26***	0.25***
	(0.01)	(0.02)
Observations	12 214	12 211

[1] Three missing cases in the location measure. Model 2 with fixed effects at 178 geographical locations
***p < .01. ** p < .05. * p < .10.

Based on the above analyses, the algorithmic recruitment process at FoodMarket seems to enhance already existing patterns, in this case the domination of younger women with European sounding surnames. These findings are clearly in line with the more dystopic narrative on AI implementation in the recruitment process. Since we observe ethnic and age discrimination, and not gender discrimination in employment in Sweden (Bursell 2014; Bygren, Erlandsson and Gähler 2017), the algorithmic recruitment studied here seems to disadvantage two already disadvantaged categories; job applicants who are older or who have non-European sounding names. Why, then, is this the case? The pessimistic narrative suggests that algorithms tend to be biased. However, it would be premature, at this point, to conclude that the algorithmic process at FoodMarket is biased. This algorithm sorts the applicants based on qualifications and test scores, and to the extent that the test scores reflect productivity, the observed outcome may simply reflect meritocracy. Since test scores are not a component of traditional recruitment at FoodMarket, it is not possible to control for test scores in the comparison of traditional and algorithmic recruitment. We can, however, assess whether sex, age and geographical background predict test scores in the algorithmic recruitment process.

As evident from Figure 1, there are mean difference in test score for the three categories in focus in this study. Women and younger employees have a much higher mean test score than men and older employees. However, employees with non-European sounding surnames have somewhat *higher* scores than employees with European-sounding names.

Figure 1. Mean test score by category among new employees at FoodMarket



In Table 3, we report results from OLS regressions assessing whether these three demographic variables predict test scores in the algorithmic recruitment process. In both model 1 and model 2, women are associated with 4.2 higher test scores, and applicants of an age below 40 are associated with almost 5 points higher test scores. Having a European-sounding name is associated with having 1.9 points *lower* test scores than for having a non-European sounding name. Thus, test scores may explain why women and younger applicants are more likely to be hired with algorithmic recruitment, but not why applicants with European sounding names are more likely to be recruited through this process. Thus, to the extent that test scores is a fair measure of productivity - a question beyond the scope of this study to address – the lower rates of new employees with non-European sounding names recruited through the algorithmic process compared with traditional recruitment, remains unexplained.

Table 3. OLS-regression predicting test score performance, standard errors in parenthesis

	Model 1	Model 2
European sounding name	-1.89***	-1.90***
	(0.43)	(0.43)
Female	4.21***	4.21***
	(0.34)	(0.34)
Below forty	4.96***	4.96***
	(0.61)	(0.62)
Position	NO	YES
Constant	73.10***	73.10***
	(0.72)	(0.72)
Observations	5615	5615

Model 2 with fixed effects at geographical locations
***p < .01. ** p < .05. * p < .10.

A closer look at *when* in the recruitment process applicants are rejected, may provide a clue to why algorithmic recruitment enhances already existing inequality. We thus turn our attention to the second data set at our disposal, the job applications sent to job openings where the algorithmic recruitment process have been employed. As explained earlier, there is no measure of age in this data set.

In Table 4, we report the share of job applicants with European-sounding and female-sounding names that have been rejected at the different stages of the job application process. It also displays share of job applicants across positions and years. Female-sounding first names make up 58 percent of all applications. They are somewhat underrepresented among those who are rejected at the earlier stages of the recruitment process, and overrepresented at the later stages. This is anticipated, since we learnt from the analyses of Table 3 that women have higher test scores.

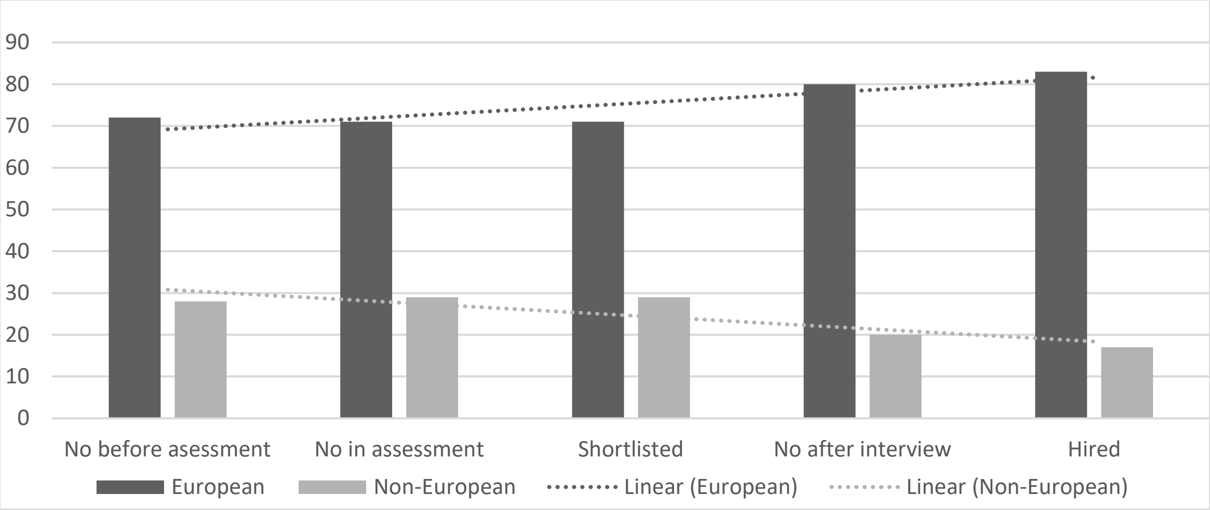
72 percent of the job applications have been submitted by individuals with European sounding surnames. Applicants with European sounding surnames are proportionally

represented among those rejected in the early stages of the recruitment process, in “Rejected before assessment”, and “Rejected in assessment”, i.e., in the algorithmic driven phases of the recruitment process. Thus, at first glance, it seems that the algorithm has selected a proportional share of European sounding names to the shortlist phase, from which managers select applications for interview. However, when the recruiting manager takes over the process from the algorithm (i.e., in “Rejected in interview” and “Hired”) applicants with European sounding names are overrepresented. Among those who are selected to interview but ultimately rejected, 80 percent have European sounding names, and among those who are hired, as many as 82 percent have European sounding names. Thus, at first glance, it seems as though the algorithm is not responsible for job applicants with European-sounding names being hired more often. Instead, it seems as though it is the recruiting managers who disproportionally selects applicants with European-sounding names from the shortlist.

Table 4. Descriptives, mean shares of European- and female-sounding names by application progress, occupation and year

Application process progress	Share European sounding	Share female
All applications	0.72	0.58
Rejected before assessment	0.72	0.51
Rejected in assessment	0.71	0.51
Rejected in shortlist	0.71	0.61
Rejected after interview	0.80	0.62
Hired	0.83	0.65
Occupations		
Store personnel	0.72	0.60
Warehouse workers	0.62	0.32
Chauffeurs	0.61	0.12
Managers	0.77	0.46
High-skilled personnel	0.80	0.61
Other	0.81	0.57
Fulltime position	0.72	0.48
Year of recruitment		
2019	0.71	0.71
2020	0.72	0.73
N	216 929	216 929

Figure 2. Application-process ladder. By European and non-European sounding names.



Since it seems to be when recruiting managers enter the recruitment process that applicants with European sounding names become advantaged, we analyze, in Table 5, the probability of reaching the interview stage for applicants that has scored high enough to be included in the shortlist. We also present models separated by whether the job vacancy is for a full-time or part-time position. In model 1 and model 2, we observe a slight advantage for European sounding names. Here, a European sounding name is associated with a 3-percentage point higher probability of reaching the interview stage. As expected, based on the descriptive information in Table 4/Figure 2, the effect of having a female-sounding names is very small at this stage of the recruitment process. In model 2, we include controls for test score, and position. Here, we see a slight decrease in the coefficient for European sounding names. The effect for female-sounding names vanishes completely with the control for test scores, just like in the models of Table 3. For full-time positions, the positive effect of having a European sounding name, is larger, 6 percentage points in model 3, and 5 percentage points when controlling for test scores and position. Thus, applicants with non-European sounding names are more disadvantaged when it comes to the more attractive fulltime positions.

Table 5. LPM models on probabilities of reaching interview from shortlist, standard errors in parenthesis

	Part-time positions		Full-time positions	
	Model 1	Model 2	Model 3	Model 4
European sounding name	0.03***	0.03***	0.06***	0.05***
Female	0.01***	0.01***	-0.02***	0.005
Test scores		YES		YES
Position		YES		YES
Constant	-0.04***	-0.05***	0.11***	0.13***
Observations	90 758	90 758	14 370	14 370

***p < .01. ** p < .05. * p < .10.

The analysis above thus clearly indicates a peculiar measure of meta-algorithmic selection bias in the final parts of the recruitment process, the parts that are handed over by the HR personnel to the individual recruiting managers.

Concluding discussion

Summing up, our analysis of FoodMarket's employment records showed that algorithmic recruitment resulted in less diversity among new employees compared with traditional recruitment. Our analysis showed that it resulted in fewer men, fewer older employees, and fewer employees with non-European sounding surnames being hired by the organization. Compared with the traditional recruitment system, women were 6 percentage points more likely to be hired; the corresponding figure for candidates with European sounding names 7 percentage points more likely, and employees below 40 years 11 percentage points more likely, to be hired with the algorithm-based evaluation process. Test scores only explained this change for women and younger applicants; the analysis of the job applications was therefore focused on the latter category. It showed that applicants with non-European-sounding names are more disadvantaged in employment to fulltime positions and that it is the recruiting managers, not the algorithm, that disproportionally rejects these applicants.

This supports the view that there is a special type of selection bias occurring in the process, despite the work being done by the algorithm; that is, the recruiting managers were responsible for the lower diversity in the hiring process, which may seem puzzling, since traditional recruitment generated a more diverse set of employees. Why would recruiting managers recruit a more diverse set of employees when they control the entire process, compared with when the first part of the process is delegated to an algorithm?

Our findings speak to the literature on human-machine interaction effects; the way people go about their job tasks changes when some of these tasks are delegated to AI and algorithms (Krakowski et al. 2019; see also Jarrahi et al. 2021). In trying to grasp this relational nexus we introduced the notion of meta-algorithmic judgment, to shed light on the algorithm as folded in the evaluation process, and to pinpoint the decision situation that confronts the recruiting manager. To illustrate, consider the case where a person usually conducts Task 1, and based on Task 1, she conducts Task 2. Delegating Task 1 to an algorithm facilitates her work process, but the algorithm provides her with an output that she knows much less about than if she had conducted Task 1 herself. Regardless of whether she believes that the algorithm has done the work properly or not, she will have to rely on it and execute Task 2 based on the algorithms' output. This fact may, in some cases, change *how* she conducts Task 2. Against this backdrop, the fact that the recruiting managers employ a different type of employees when interacting with a machine, is not puzzling per se.

But why do the recruiting managers select more applicants with European sounding names when the algorithm provides a competence-based shortlist? About this, we can only speculate, yet these speculations can be systematically evaluated in the future and help us make progress when theorizing the results from new empirical studies (Lave and March 1975; Swedberg 2021). One possible explanation is, that recruiting managers become more risk averse when they lack the in-depth knowledge about the applicants in the shortlist. Previous research has shown that people become more prejudiced and discriminate more when they lack information on the people they are evaluating (Kunda and Thagard 1996). It can also be the case, that if the managers feel that the recruitment algorithm are decontextualizing and eliminates biases in a way that misses what they believe is really needed in the day-to-day activities, then they might start improvising and use their own professional strategies. Another possibility is that the managers, who have been informed that the algorithm is unbiased, have implicitly delegated the responsibility for diversity to the algorithm, and somehow let go of the responsibility for diversity in their own decisions.[7] A final possibility is that in traditional recruitment, managers hire a more diverse set of employees because they prefer another "type" of non-European applicants than those that are selected based on test scores.[8] They might, for instance, prefer non-European applicants who signal "Swedishness": secular values, or assimilation signaled in the personal letters or through appealing photographs. When the application process is stripped of these features, managers may settle, to a higher extent, for what they perceive as in-group candidates.

While the higher share of women and younger applicants can be viewed as the result of a meritocratic process, it could also, potentially, be a case of indirect discrimination, if the questions in the work test somehow systematically disadvantage men and older applicants. Answering this question would require an in-depth study of the test and why men and older applicants score lower, a task beyond the scope of this study to address. However, regardless of the potential outcome of such a study, it is, from a societal perspective, important to supervise these types of effects. FoodMarket, will, with a yearly staff turnover of around 13 percent (a standard turnover rate in this industry), change the demographic compositions of

[7] A similar phenomenon seems to have occurred when an AI-system was set to provide recommendations about unemployed clients at the Employment Office in Poland. The employment officers followed the recommendations to such a high extend that the algorithmic recommendations in practice became decisions (Misuraca and van Noordt 2020).

[8] Informants at the HR department has confirmed that the some recruiting managers did express discontent with the applicants included in the algorithms' shortlists.

their employees in line with patterns identified in this study in only a few years' time if they continue to recruit with the current algorithmic process.

When it comes to FoodMarket, it stands out as an employer who is strongly committed to its goals of diversity and equal treatment. There is certainly a selection bias in what type of firms that share their data with external researchers; FoodMarket have generously shared their data and given us of their time, assisting us in understanding their work processes and data, and they have taken a genuine interest in the results of our evaluations. We are confident that they will continue to improve and evaluate their recruitment processes and make efforts to mitigate unforeseen consequences. However, all organizations that automate their recruitment processes will likely neither have the resources nor the commitment to make similar efforts. Regardless of why men and older applicants score lower, or why recruiting managers employ a less diverse set of employees with the support from a recruitment algorithm, our study show that this technological shift may create new groups of winners and losers, an issue that must be continuously scrutinized when an increasing number of public and private organizations starts making crucial decisions based on algorithms. These decisions will influence people's life chances, which is reason enough for maintain a sharp focus on this socio-tecnological development in the future.

References

Ajunwa, I. 2020. “The paradox of automation as anti-bias intervention.” *Cardozo Law Review*, 41(5): 1671-1742.

Baert, S. 2018. “Discrimination in the labour market: A register of (almost) all correspondence experiments since 2005”, pp. 63-77 in S. M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. Springer.

Bader, V. and Kaiser, S. 2019. “Algorithmic decision-making? The user interface and its role for human involvement in decisions supported by artificial intelligence.” *Organization* 26(5): 655-672.

Bareis, J. and Katzenbach, C. 2021. “Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics.” *Science, Technology, & Human Values* 1-27.

Beer, D. 2017. “The Social Power of Algorithms.” *Information, Communication & Society,* 20(1): 1-13.

Bovenkerk, F. 1992. *Testing Discrimination in Natural Experiments: a manual for international comparative research on discrimination on the grounds of ‘race’ and ethnic origin*. Geneva: ILO.

Burrell, J. 2016. “How the machine ‘thinks’: Understanding opacity in machine learning algorithms.” *Big Data & Society* 1-12.

Burrell, J. and Fourcade, M. 2021. “The Society of Algorithm.” *Annual Review of Sociology* 47: 213-237.

Bursell, M. 2014. “The Multiple Burdens of foreign-named men: Evidence from a field experiment on gendered ethnic hiring discrimination in Sweden.” *European Sociological Review* 30(3): 399-409.

Bursell, M. Bygren, M. and Gähler, M. 2021. “Does Employer Discrimination contribute to the subordinate labor market inclusion of individuals of a foreign background?” *Social Science Research* 98: 102582.

Bursell M. and Olsson F. 2021. “Do we need dual-process theory to understand implicit bias? A study of the nature of implicit bias against Muslims“. *Poetics - Journal of Empirical Research on Culture, the Media and the Arts*. Published online March 16. https://www.sciencedirect.com/science/article/pii/S0304422X21000267

Bursell M. and Olsson F. 2020. “On the difficulty of identifying implicit racial beliefs and stereotypes.” *American Sociological Review* 85(6):1117-1122.

Diakopoulos, N. 2016. “Accountability in Algorithmic Decision Making.” *Communication of the ACM* 59(2): 56-62.

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

van Esch, P., Black, J. & Arli, D. 2020. “Job candidates’ reactions to AI-Enabled job application processes.” *AI Ethics*. https://doi.org/10.1007/s43681-020-00025-0

Misuraca, G., and van Noordt, C. 2020. *Overview of the use and impact of AI in public services in the EU*, EUR 30255 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19540-5, doi:10.2760/039619, JRC120399.

Fazi, MB. 2021. “Beyond Human: Deep Learning, Explainability and Representation.” *Theory, Culture, & Society* 8(8): 88-88.

Feenberg, A. 2002. *Transforming Technology: A Critical Theory Revisited*. Oxford University Press.

Glaser, VL. Pollock, N. and D’Adderio L. 2021. “The Biography of an Algorithm: Performing algorithmic technologies in organizations.” *Organization Theory* 2:1-27.

Greenwald, A. and M. Banaji. 1995. “Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes.” *Psychological Review* 102:4–27.

Howard, A. and Borenstein, J. 2018. “The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity.” *Science and Engineering Ethics* 24: 1521-1536.

Jarrahi, MH, Newlands G, Lee MK et al. 2021. “Algorithmic management in a work context.” *Big Data & Society* 1-14.

Jaton, F. 2021. “Assessing biases, relaxing moralism: On ground-truthing practices in machine learning design and application.” *Big Data & Society* 1-15.

Kahneman, D. Sibony, O. Sunstein, CR. 2021. *Noise: A Flaw in Human Judgment*. New York: Little Brown Spark & Co.

Keeley, B. 2015. *Income Inequality: The Gap between Rich and Poor*, OECD Insights, OECD Publishing, Paris, https://doi.org/10.1787/9789264246010-en.

Kiviat, B. 2019. “The moral limits of predictive practices: The case of credit-based insurance scores.” *American Sociological Review*, 84(6): 1134–1158.

Kleinberg J, Ludwig J, Mullainathan S, and Sunstein CR. 2019. “Discrimination in the age of algorithms.” *Journal of Legal Analysis* 10: 113-174.

Krakowski, S., D. Haftor, J, Luger, N. Pashkevich, S. Raisch. 2019. “Humans and Algorithms in Organizational Decision Making: Evidence from a Field Experiment.” *Academy of Management*. Published online 1 Aug 2019. https://doi.org/10.5465/AMBPP.2019.16633abstract

Krogh von, G. 2018. “Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing.” *Academy of Management Discoveries* 4(4): 404-409.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*(2), 284–308.

Köchling, A. and Wehner, MC. 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: 795-848.

Larson, EJ. 2021. *The Myth of Artificial Intelligence: Why Computers Can’t Think the Way We Do*. Harvard, MA: Harvard University Press.

Lauer, J. 2017. “End of judgment: Consumer credit scoring and managerial resistance to the black boxing of creditworthiness.” In Raff, D. M. G., Scranton, P. (Eds.), *The emergence of routines: Entrepreneurship, organization, and business history*. Oxford: Oxford University Press.

Lave, CA. and March, JG. 1975. *An Introduction to Models in the Social Sciences*. New York: Harper & Row

Lee, F. Bier J. Christensen j. et al. 2019. “Algorithms as folding: Reframing the analytical focus.” *Big Data & Society* 6(2): 1-12.

Logg, JM. Minson, JA. and Moore, DA. 2019. “Algorithm appreciation: People prefer algorithmic to human judgment.” *Organizational Behavior and Human Decision Processes* 151: 90-103.

Makridakis, S. 2017. “The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms.” *Futures* 90:47-60.

Moss, PM, and Tilly, CT. 2001. *Stories Employers Tell. Race, Skill, and Hiring in America*. Russel Sage, Chicago.

March, JG. 1978. “Bounded rationality, ambiguity, and the engineering of choice.” *Bell Journal of Economics*, 9: 587–608.

McCarthy, J. 1979. “Ascribing mental qualities to machines.” In: *Philosophical perspectives in artificial intelligence* (ed.) M. Ringle, Atlantic Highlands, NJ: Humanities Press.

Newman, DT. Fast, NJ. and Harmon, DJ. 2020. ”When eliminating bias isn’t fair: Algorithmic reductionism and procedural justice in human resource decisions.” *Organizational Behavior and Human Decision Processes* 160: 149-167.

Noble, S. 2018. *Algorithms of Oppression: How Search Engines Reinforces Racism*. New York: New York University Press.

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.

O’Neil, C. 2016. *Weapons of Math Destructions: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin.

Pager, D., B. Bonikowski and B. Western. 2009. Discrimination in a Low-Wage Labor Market: A Field Experiment *American Sociological Review* 74: 5:777-99.

Polack, P. 2020. “Beyond algorithmic reformism: Forward engineering the design of algorithmic systems.” *Big Data & Society* jan-jun: 1-15.

Raghavan, M. Barocas, S, Kleinberg, J. Levy, K. 2019. “Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices.” *arXiv* arXiv:1906.09208.

Ridgeway, C. 2019. *Status. Why Does It Matter? Why Is It Everywhere?* New York: Russell Sage.

Rivera, LA. 2015. “Go with Your Gut: Emotions and Evaluation in Job Inter-views.” *American Journal of Sociology* 120(5):1339-1389.

Roumbanis, L. 2017. “Academic judgments under uncertainty: A study of collective anchoring effects in Swedish Research Council panel groups.” *Social Studies of Science* 47(1): 95–116.

Roumbanis, L. 2022. “Disagreement and Agonistic Chance in Peer Review.” *Science, Technology, & Human Values*. Online first: doi.org/10.1177/01622439211026016.

Russell, S. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking Press.

Schiølin, K. 2019. “Revolutionary dreams: Future essentialism and the sociotechnical imaginary of the fourth industrial revolution in Denmark.” *Social Studies of Science* 50(4): 542-566.

Searle, JR. 1980. “Minds, brains, and programs.” *Behavioral and Brain Sciences* 3(3): 417-457.

Sloman, A. 2014. “How can we reduce the gulf between artificial and natural intelligence?” *International workshop on Artificial Intelligence and Cognition* (AIC 2014), November 26-27, University of Turin.

Stilgoe, J. 2018. “Machine learning, social learning and the governance of self-driving cars.” *Social Studies of Science* 48(1): 25-56.

Sunstein, C. 2019. “Algorithms, Correcting Biases.” *Social Research* 86(2): 499-511.

Swedberg, R. 2021. “Does Speculation Belong in Social Science Research?” *Sociological Methods & Research* 50(1): 45-74.

Turing, A. 1950. “Computing machinery and intelligence.” *Mind* 59(236): 433-460.

Upadhyay, A.K. & K. Khandelwal. 2018. “Applying artificial intelligence: implications for recruitment.” *Strategic HR Review.* 17:5:255-258.

Vrontis, D., M. Christofi, V. Pereira, S. Tarba, A. Makrides & E. Trichina. 2021. Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review, *The International Journal of Human Resource Management*, DOI: 10.1080/09585192.2020.1871398

Winner, L. 1980. “Do Artifacts have Politics?” *Daedalus* 109(1): 121-136.

Yeung, K. 2018. “Algorithmic regulation: A critical interrogation.” *Regulation & Governance*, 12(4): 505–523.

Yu, M. 2021. “The Algorithm Concept, 1684-1958.” *Critical Inquiry*, 47(3): 592–609.

Zarsky, T. 2016. “The Trouble with Algorithmic Decisions: An Analytical Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making.” *Science, Technology, & Human Values* 41(1): 118-132.

Ziewits, M. 2016. “Governing Algorithms: Myth, Mess, and Methods.” *Science, Technology, & Human Values* 41(1): 3-16.

Åslund, O. & O. Nordström Skans. 2012. “Do Anonymous Job Application Procedures Level the Playing Field?” *ILR Review*, 65:1:82–107.

Appendix A

Name classification first names

For people with multiple space-separated names, only the first name that is in the database was used (*this was to avoid confusion due to the fact that many non-Swedish double names seem to be gender-mixed, e.g., “Jorge María”). If a person’s given name(s) was/were not in the database, a first run over the data was done to check whether or not there were any names within a Levenshtein distance of 1 from the original name in the SCB database. If those names exist, the ratio of the total number of women with those names divided by the total number of people with those names was used as the probability estimate. If a person’s given name(s) was/were not in the database and had no neighbours of Levenshtein distance 1, or simply had no given name registered, no probability was assigned. All names have, prior to gender probability estimation, been transformed as to remove non-alphabetical characters (such as apostrophes) except for spaces, which were preserved (see above on how double names have been handled), and dashes, which were replaced by spaces. The names were also transformed into all lower-case letters.

Name classification surnames

The classifier model was trained on data from biographical articles of living people on the English language Wikipedia. The names of the subjects were matched to a country of origin based on the article’s belongingness to geographical categories (categories like e.g. “People from Stockholm, Sweden”). Countries around the world were then placed in nine categories – Scandinavian, Finnic, Western European and North American, Eastern European, Southern European, Middle eastern/Muslim, African non-Muslim, Asian non-Muslim and Latin American. The assumption is that the types of names in each of these geographical regions will not overlap to much with those in the other regions.

All data was transformed in the following manner: all non-alphabetic characters except for dashes were deleted. All language-specific Latin characters were converted to their nearest (by looks) English equivalent (e.g., the Vietnamese “đ” were converted to lower case d). All capital letters were converted to lower case. For observations with multiple last names, only the final last name was used. The specific model used for the classification was a neural network consisting of three stacked bidirectional LSTM layers.