Fairness, AI & recruitment

Rigotti, C.; Fosch Villaronga, E.

Computer Law & Security Review 53 (2024) 105966

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

journal homepage: www.elsevier.com/locate/clsr

Fairness, AI & recruitment

Carlotta Rigotti [*], Eduard Fosch-Villaronga

eLaw Center for Law and Digital Technologies, Leiden University, the Netherlands

ARTICLE INFO

Keywords:
Fairness
Hiring
Recruitment
Discrimination
Data protection
Artificial intelligence

ABSTRACT

The ever-increasing adoption of AI technologies in the hiring landscape to enhance human resources efficiency raises questions about algorithmic decision-making's implications in employment, especially for job applicants, including those at higher risk of social discrimination. Among other concepts, such as transparency and accountability, fairness has become crucial in AI recruitment debates due to the potential reproduction of bias and discrimination that can disproportionately affect certain vulnerable groups. However, the ideals and ambitions of fairness may signify different meanings to various stakeholders. Conceptualizing fairness is critical because it may provide a clear benchmark for evaluating and mitigating biases, ensuring that AI systems do not perpetuate existing imbalances and promote, in this case, equitable opportunities for all candidates in the job market. To this end, in this article, we conduct a scoping literature review on fairness in AI applications for recruitment and selection purposes, with special emphasis on its definition, categorization, and practical implementation. We start by explaining how AI applications have been increasingly used in the hiring process, especially to increase the efficiency of the HR team. We then move to the limitations of this technological innovation, which is known to be at high risk of privacy violations and social discrimination. Against this backdrop, we focus on defining and operationalizing fairness in AI applications for recruitment and selection purposes through cross-disciplinary lenses. Although the applicable legal frameworks and some research currently address the issue piecemeal, we observe and welcome the emergence of some cross-disciplinary efforts aimed at tackling this multifaceted challenge. We conclude the article with some brief recommendations to guide and shape future research and action on the fairness of AI applications in the hiring process for the better.

1. Introduction

In an era where technology reshapes industries and the way we work [1], the integration of artificial intelligence (AI) in recruitment processes has emerged as a significant trend [2–6]. Because this technology can process data and make decisions at volumes and speeds far surpassing human capabilities, it enhances the efficiency in identifying, attracting, screening, evaluating, interviewing, and managing job applicants [7]. However, AI applications for recruitment and selection purposes are not flawless and can reproduce and perpetuate diversity bias, thereby discriminating against job applicants because of their personal characteristics [8,9]. The Amazon resume selection algorithm is a well-known example, with its design relying on historical gender imbalances in the company's hiring practices, favoring men over women for technical roles [10]. Although the Amazon resume selection algorithm was never launched on the market, it still offers a practical illustration of the substantial, detrimental effects that can arise from neglecting diversity and inclusion in training algorithms. If deployed without careful consideration, the potential repercussions on the job applicant's dignity, autonomy, and well-being extend to various facets of their life. This includes but is not limited to, adverse effects on social participation, economic circumstances, housing opportunities, family dynamics, and potential impacts on physical and mental health. Consequently, if human resources (HR) practitioners and employers plan to increasingly benefit from the advantages of technological innovation, they must understand whether using these tools could lead to hidden negative repercussions and be empowered to mitigate the related risks. Said otherwise, HR practitioners and employers must ensure that their AI applications uphold *fairness* in the hiring process [11].

As with many other abstract concepts that serve the goals of the law, nevertheless, we are faced with the challenge of how we should define and operationalize the concept of fairness in the context of AI applications for recruitment and selection purposes, which strikingly has become an elusive concept with no clear, accepted definition. Indeed,

* Corresponding author.
E-mail address: c.rigotti@law.leidenuniv.nl (C. Rigotti).

https://doi.org/10.1016/j.clsr.2024.105966

Available online 8 April 2024

although this research question might appear straightforward, fairness has become an increasingly appealing yet remarkably nebulous concept, with legislators and scholars from different disciplines frequently using it without a precise and common definition [12]. However, not having a precise legal definition for fairness blurs what counts as a legal or illegal practice, a narrative used more often than it should in the AI discourse to support 'the need for ethical principles and to exclude legal rules as if they were equal or interchangeable' [13]. In this respect, we claim that what qualifies as fair in AI applications in the hiring process requires more precise delineation to ensure legal certainty concerning the roles and responsibilities of the ecosystem surrounding the creation of these tools and their further application in these processes, and the guarantee of the user rights in an increasingly automated workplace [14].

This article aims to contribute to the literature by delineating fairness within AI-driven selection and recruitment, encompassing all HR practices from creating job vacancies to signing employment contracts. We do so through a scoping literature review at the intersection of labor, legal, and other AI studies in this context. In doing so, we assume that fairness constitutes situated knowledge and practices [15] and acknowledge that it is not a one-size-fits-all concept but is contingent on specific contexts. We interchangeably use "the recruitment and selection process" and "the hiring process" in this discussion.

We have organized this article in the following way. Section 2 briefly provides a historical overview of technological innovation within the recruitment and selection process in the past decades. It also highlights its ambitions and limitations. In Section 3, we explore the concept of fairness from the perspective of the stakeholders involved in the recruitment and selection process, namely job applicants and HR practitioners. Section 4 looks at the legal landscape, explicitly examining how fairness has been defined and promoted within data protection and anti-discrimination law and the proposed regulation for AI (also called the AI Act, or AIA). While other legal frameworks may also be relevant, the focus here stems from the common recognition that AI applications for recruitment and selection process personal data may lead to discriminatory outcomes for job applicants. Furthermore, it is noteworthy that data protection, anti-discrimination, and AI laws have generated a richer and rapidly expanding literature in the realm of fairness and AI applications in the labor market. Subsequently, since, in theory, academic independence should lay the theoretical and methodological foundations and cross-disciplinary efforts should bridge the gap between theory and practice, shaping policies and standards guiding the responsible development and deployment of technology [16], Section 5 discusses whether the definition and respect for fairness arises from independent work or is instead the result of a cross-disciplinary research and policy effort. While recognizing the need to fill terminological gaps and harmonize competing understandings, Section 6 includes some brief propositions and food for thought to guide future research and action in upholding the fairness of AI applications for recruitment and selection purposes. In Section 7, we summarize the findings of this research article.

2. Introducing AI applications in the recruitment and selection process

Before the late 1990s, the recruitment and selection processes relied on manual methods. HR practitioners were responsible for attracting job seekers, manually screening and assessing job applications, and determining who should advance in the hiring process or secure employment. However, the traditional nature of this recruitment and selection approach proved to be labor intensive, and it often exposed deliberate and unintentional biases from HR practitioners [3,17,18]. This means that professionals involved in the hiring process frequently engage in conscious or unconscious stereotyping and discrimination against job applications due to their personal characteristics, such as gender and age.

However, the hiring process underwent a significant transformation with the advent and swift proliferation of the Internet during the 1990s. One notable event was the advent of digital job boards, which began to gather and post several job vacancies, targeting and reaching a vast pool of potential job candidates at minimal expense and in a more attractive manner [19]. In this context, the network effect played a pivotal role: as websites showcased more job vacancies, they drew in more job seekers, and as they attracted more job seekers, they encouraged employers to post more job vacancies and pay for their services [20]. Simultaneously, online recruitment (or e-recruitment) took the shape of professional networking platforms, meaning that people could form and cultivate a community centered around work-related interests, facilitating the exchange of information and endorsements[3,17,18]. Platforms such as LinkedIn, Glassdoor, Indeed, or Monster, among many others [21,22], clearly illustrate this point. Over time, this wave of job seekers encouraged more employers to post job vacancies on these platforms, creating a self-reinforcing cycle of growth and connectivity, which gradually paved the way for further advancements in AI-powered solutions.

To date, HR practitioners navigate an ever-evolving landscape, often striving to keep up with technological advancements [23]. Increasingly, however, they recognize AI's transformative potential in talent acquisition and are turning to AI-driven methods for recruitment and selection [21,22,24].

2.1. The ambitions of AI-driven recruitment and selection

This transition represents a plot twist in how organizations hire new personnel, specifically the development and use of AI applications in the hiring process that simplify or replace HR practitioners in executing four key functions: outreach, screening, assessment, and coordination [3,8,25]. This implies that AI applications currently have the potential to aid HR practitioners in the attraction and identification of job applicants because they can easily learn and strategically place job vacancies through various means, such as banners, pop-ups, emails, and text messages, to maximize visibility and responses (something used in digital marketing, see [26]). In this sense, AI applications can outperform human counterparts in screening job applications thanks to their ability to expedite the process and extract specific skills and personality traits from a job applicant's digital records, including their online presence on social media platforms [27]. They can also enhance HR practitioners in the assessment phase, with gamification as a classic example, and streamline the coordination across different stages of the recruitment and selection process [3,28,29]. Throughout this hiring pipeline, decisions at each stage generate data likely to influence the following interactions, forming a feedback loop. For instance, assessment outcomes can impact job tenure, a key prediction target for future outreach and screening driven by AI applications [23].

Although the development and use of AI applications attempt to simplify or sometimes substitute HR practitioners, the increasing levels of automation do not lead to the direct exclusion of human beings from the process. As it already happens in surgery, civil aviation, and the board of directors automation [30], human involvement remains indispensable. HR practitioners are responsible for using and overseeing the technology to ensure it operates within specified parameters, with the recruitment and selection process facilitated by the machines [31]. In this context, it is argued that AI applications lack the intuitive, emotional, and content-sensitive capabilities inherent to human beings, which are crucial for effective functioning [32]. Furthermore, human engagement always remains during the design phase, which includes developing the technology, determining which datasets are utilized, and supervising the training and testing phases [28].

Generally, the recruitment and selection of the right people are inextricably intertwined with the survival and prosperity of any organization. Said otherwise, the hiring process creates a pool of workers possessing the optimal blend of knowledge, skills, abilities, and other attributes necessary for gaining a competitive edge [3,33,34]. In this regard, adopting AI applications for recruitment and selection stems from a strategic motivation driven by the so-called 'war for talent' [35]. The prevailing belief is that AI applications can enhance the efficiency of virtually any recruitment and selection process, particularly in terms of time, cost, and effort [[28,36–39]. In the conventional hiring paradigm, time and geography usually emerge as critical limitations since the recruitment process can be time-consuming, characterized by lengthy job postings, multiple rounds of interviews, and complex decision-making procedures often bound to geographical restrictions. With AI-driven tools, organizations can surpass the constraints of geographical boundaries, enabling them to reach a global talent candidate pool without being bound by physical proximity. Using AI applications could introduce greater flexibility into the hiring process by eliminating constraints related to time and location [28,37], and dramatically expedite this process, by streamlining various aspects of recruitment. For instance, automated screening algorithms can swiftly sift through vast pools of resumes, identifying the most qualified candidates in a fraction of the time it would take a human recruiter. Chatbots and virtual assistants can engage with candidates around the clock, assuring that interactions and initial assessments are flexible, adapt to candidates' needs, and are not confined to business hours [40]. Furthermore, AI-driven solutions can reduce the administrative burden associated with interview scheduling and coordination, often a challenging and not necessarily fulfilling part of the recruitment process [41]. Automating these tasks can make the overall process more efficient and responsive to employers and candidates.

In addition to these advantages, technological innovation can assist HR practitioners in striking a balance between their intuitive and creative decision-making abilities and the analytical capacity AI offers in handling complex and numerous data [42]. For instance, predictive analytics and machine learning models could help anticipate talent needs in advance, allowing organizations to maintain talent pipelines and proactively identify candidates even before specific roles are open [40,43]. Furthermore, AI applications seem to excel in identifying the most crucial (and sometimes less evident) criteria for matching job applicants with job vacancies [34]. Finally, since AI applications appear to execute recruitment and selection tasks without immediate human involvement, there is a belief among some scholars that their decisions are less biased and better ensure fairness [3,36,44].

2.2. *The limitations of AI-driven recruitment and selection*

Against these technological promises, a growing body of literature raises concerns about adopting AI applications within the recruitment and selection process. Notably, the creation and implementation of this technology come with considerable costs, creating a situation where large corporations maintain a competitive advantage while small and medium-sized enterprises struggle to harness its benefits [3]. Furthermore, for AI applications to function effectively, they must handle personal and sensitive data, which in turn could impinge on the privacy and data protection of job applicants [2], at risk of their social marginalization, stigmatization, and discrimination too [3,45,46]. Apart from potentially leading to practices that may compromise fundamental rights, this is also significant because AI applications work in a way that is likely to make decisions complex to be comprehended fully by humans [28,47]. Consequently, the technology might lack transparency and explicability, undermining accountability and responsibility [12,38]. Because of these concerns, some HR practitioners are skeptical about the adoption of AI applications for recruitment and selection purposes due to concerns that they pose a threat to their jobs rather than serving as supportive tools [3,48,49].

When trying to comprehensively examine the adoption of AI applications in the recruitment and selection process, some research focuses on how job applicants react. This entails investigating the emotions, attitudes, and behaviors arising from their interactions with AI-driven recruitment and selection methods. Within this domain, empirical research has yielded contrasting findings regarding job applicants' perceptions of fairness and diversity bias when engaging with AI applications. On the one hand, some scholars hold an optimistic view of job applicants' potential receptiveness to AI applications. Job applicants appear to associate the use of AI applications with innovation, which, in turn, enhances their attraction to organizations [50]. Conversely, others contend that job applicants regard AI-driven recruitment and selection processes less equitable [51]. For instance, in a recent study surveying some candidates' perceptions of AI-driven evaluations and their perception of justice, participants stressed their preference for keeping humans in the evaluation process loop [52]. To that respect, despite acknowledging that humans have inherent biases, the interviewed participants preferred them because those were the 'devil they knew' instead of the 'unknown devil' represented by the algorithm. Other studies indicate that job applicants express dissatisfaction with the impersonal nature inherent in AI-driven recruitment and selection processes, which can demotivate them from applying to new job postings [33,28].

Another adverse perception among job applicants of integrating AI applications in recruitment and selection processes is their unfamiliarity with the technology. This sentiment may be closely intertwined with the perceived lack of transparency in how the technology is designed, its inner workings, and its deployment. Transparency here refers to providing information about the recruitment and selection process, making it more predictable and justifiable from the job applicants' perspective [28,37,52]. Privacy and data protection concerns, as well as apprehensions about potential discrimination, are also voiced by job applicants [33,37,53]. Some scholars point out that this technology is a double-edged sword in the context of professional network platforms, which are frequently AI-driven, such as LinkedIn. While it can enhance job seekers' visibility and opportunities, it exposes them to new vulnerabilities. It restricts individuals from customizing the extent and nature of disclosed information depending on the audience, context, or level of established trust [22].

In summary, AI applications are progressively designed and used to carry out several tasks related to the recruitment and selection process. However, this design and use come with aspirations from deployers, companies, and enthusiastic early adopters and constraints upon the verge of infringing fundamental rights. Beyond the desire for innovation and improved efficiency, there is a prevalent concern regarding potential rights violations and diversity biases that applicants also perceive. With this in mind, the following section explores what could constitute fairness in the context of AI applications within the hiring process.

3. Defining AI-driven fairness in the recruitment and selection process: The stakeholders' perspectives

The previous section demonstrated that technological innovation has increasingly driven the recruitment and selection process to support HR practitioners and enhance efficiency. Nevertheless, it also showed that this pursuit comes with various constraints that can significantly contribute to the perceived unfairness of AI applications in the hiring process, especially regarding privacy violations and social discrimination against job applications. However, these constraints vary in significance based on one's role in the recruiting process, making fairness a subjective concept fueled and influenced by individual experiences and expectations, which diverge for job applicants and HR officers. Upon closer examination, matters concerning perceptions of violations become more intricate, with this section revealing that fairness is inherently multifaceted and contingent on one's position within the recruitment and selection process. In other words, perceptions of fairness in the recruitment and selection process largely arise from different stakeholders' perspectives. A growing body of scholarly work underscores the dynamic nature of fairness and posits that the perception of fairness evolves throughout the distinct stages of the selection and recruitment process [54–56]. This perspective demonstrates that, although an overarching definition of fairness may help delineate an operating framework, it is necessary to tailor that definition to the specific phase within the hiring process.

3.1. Job applicants

Qualitative research proves that many job applicants compare their understanding of fairness with their desire for equitable treatment and outcome, in the sense that their knowledge, skills, and efforts should seemingly match with the hiring decision of the HR practitioner(s) [57–60]. To their understanding, they relate fairness to their personal experience of respect, dignity, and honesty during the hiring process, emphasizing how critical human empathy and communication are in an employment setting [61,10].

However, due to the information asymmetry inherent to most selection and recruitment processes, job applicants are usually more inclined to see fairness through procedural lenses. This means that they believe that the hiring process serves to achieve fair hiring outcomes. In the literature [52,57,62–69], this perception of fairness has been named 'procedural fairness,' with its more common principles being the following ones:

- Job relatedness: The hiring process should only evaluate that information necessary for the job post, which could, in turn, help predict the skills and capabilities of the job applicant.
- Consistency: Each job applicant should go through the same hiring process.
- Opportunity to perform: The job applicant should have the chance to prove their knowledge and skills during the hiring process.
- Objectivity: The assessment of the job applicant should rely on relevant and impartial criteria, thereby excluding personal biases or subjective opinions of the HR practitioners.

Given the procedural nature of fairness, job applicants' perceptions of fairness may hinge on whether the recruitment and selection processes are analog or AI-driven and may depend on their personal experience. Generally, AI-driven processes tend to be associated with negative attitudes [70–72] because of the job applicant's perception of having less control and influence in AI-driven processes than in traditional ones [70]. Conversely, some other studies found that individuals frequently encountering social discrimination in their workplaces believe that deploying AI applications can enhance fairness and make organizations more attractive [71].

In summary, job applicants seemingly assess fairness based on their personal experiences and some procedural fairness principles. The introduction of AI in the hiring process adds complexity, with various perceptions - some negative due to reduced control and others positive, especially for those facing discrimination. Potentially, balancing these different perspectives is key to establishing fair recruitment and selection on a social level.

3.2. HR practitioners

While some studies have tried to capture the workers' attitudes toward automation in the workplace, the literature review that we conducted revealed fewer studies dedicated to understanding the perception of fairness from the HR practitioner(s) standpoint within the hiring process.

Some studies that aimed to capture precisely this emphasize the significance of the person-job fit in this context [61]. Said otherwise, HR practitioners might construe fairness in the selection and recruitment processes concerning matching the knowledge, skills, abilities, and other personal characteristics of the job applicant and the core tasks, duties, and responsibilities of the job vacancy. This implies that fairness, from their perspective, emanates from a selection and recruitment process that primarily benefits the employer and immediate stakeholders of the organization. This consideration may outweigh concerns related to various perceptions of fairness amongst job applicants or broader social interests, like respect for diversity and inclusion [61]. In this scenario, it is conceivable that HR practitioners not only aim to identify the most suitable job applicant in terms of qualifications but also consider how their hiring decision fits within the broader organizational context [73]. Interestingly, findings from Sami Koivunen and colleagues (2019) indicate that HR practitioners acknowledge the importance of workforce diversity but often encounter practical challenges, as these objectives may clash with the practices or culture of specific teams or the organization. Consequently, variations in values can lead to divergent perceptions of what constitutes fairness among HR practitioners [74].

In conclusion, the translation of fairness perceptions into practical implementation is intricate. Job applicants, valuing personal experiences and procedural fairness, may find their criteria diverging from HR practitioners who often prioritize the employer's interests. Given that every hiring decision is seemingly normative, whether analog or AI-driven, the complexity arises. In this normative context, we find a compelling argument to prioritize the vulnerable party in the relationship, emphasizing fairness in a way that safeguards the interests of job applicants. This statement echoes most legal perspectives outlined in the next section, which seeks to define the fairness of AI-driven hiring processes.

4. Understanding fairness of AI applications for recruitment: A legal perspective

When assuming that the law has an expressive function aimed at coordinating human conduct and informing human beliefs [75,76], one could argue that existing regulations may provide a framework that helps us navigate the divergent perspectives on fairness among the stakeholders engaged in the AI-driven process of recruitment and selection. They could act as a normative foundation, shaping conduct and fostering a shared understanding of fairness within the complex landscape of AI applications for hiring purposes. Accordingly, in the following subsections, we examine various branches of laws governing, in one way or another, AI applications designed to recruit and select job applicants and investigate their definition and promotion of fairness in AI-driven hiring. In doing so, we first acknowledge that the term 'fairness' may not always be explicitly articulated in legal texts. Rather, scholars often infer and contextualize it through other legal principles and categories. As such, our analysis explores these scholarly interpretations. Second, we contextualize the laws and legal discourse within the framework of AI applications for recruitment and selection. This approach recognizes that technology is not merely the subject of regulation, but also serves as an analytical tool to highlight both the aspirations and constraints of the existing legal framework.

4.1. Anti-discrimination law

At the EU level, fair recruitment relies on the actual decision-making process but not its result [77]. In other words, the focus of anti-discrimination law in an employment context does not lie on the hiring decision but on the treatment of the job applicant in the recruitment and selection process, thereby implicitly favoring a procedural understanding of fairness. This approach emerges from the case law of the European Court of Justice (ECJ), which in 2008 ruled that public statements by an employer explicitly refusing to hire individuals of a certain ethnic origin constituted direct discrimination (Feryn, C-54/07). From 2010 onwards, it has been consistently held that setting age limits for recruitment in certain professions is generally not permissible unless fitness is essential for the job (Wolf, C-229/08 and C-341/08; Vital Pérez, C-416/13; Salaberria Sorondo, C-258/15). Similarly, religious affiliation requirements for job positions in churches or religious organizations must be subject to effective judicial review and must be necessary, objectively justified, and proportionate to the nature of the job (Egenberger, C-414/16).

On such premises, we explore whether anti-discrimination law could effectively ensure the fairness of AI applications used in hiring. This involves examining various definitions of discrimination and contemplating how fairness could be conceptualized in contrast.

Regarding the directives setting the general framework of anti-discrimination law, discrimination is alternatively conceptualized as follows:

- Direct discrimination pertains to situations where an individual is treated less favorably than another person in a similar circumstance due to personal attributes. With some exceptions [78], it is generally presumed that the development and implementation of AI applications rarely lead to direct discrimination [1,79]. For example, in the labor market, it is improbable that AI applications explicitly employ gender, race, or other legally protected factors to assign lower ratings when matching job applications with potential candidates. As we expand on below, this could be because current protected grounds do not capture many new algorithmic-induced discriminatory instances.
- Indirect discrimination encompasses scenarios in which a seemingly unbiased rule, criterion, or procedure places a person or a group of individuals at a distinct disadvantage compared to others unless there is a legitimate objective for such a rule and the means to achieve that objective are suitable and indispensable. Concerning AI applications in the labor market, this implies that an AI-driven selection and recruitment process may lead to indirect discrimination if it unjustifiably rejects job applications from a disproportionately large number of individuals based on their personal characteristics. Generally, it is believed that indirect discrimination is more likely to occur due to the design and deployment of AI applications because this technology predominantly relies on neutral criteria and methodologies [1,79]. However, recent literature may prove this assumption superfluous since technology is not neutral; on the contrary, from its conception to its deployment, conscious choices, which leave much room for desire in terms of diversity and inclusion, have steered the technology [80–82].

Beyond the current scope of anti-discrimination law and its definitions, most research on AI-driven discrimination supports the idea of 'proxy discrimination.' This term encompasses all forms of discrimination stemming from associations with protected characteristics [83–85]. In brief, proxy discrimination manifests whenever the AI application uses neutral information as a stand-in (*i.e.*, a proxy) for a prohibited ground. For instance, requesting the job applicant's address to deduce their race or ethnicity illustrates this point clearly [86]. Proxy discrimination is not a novel concept [87] and has traditionally been used by employers to circumvent anti-discrimination regulations. Simultaneously, while HR practitioners rarely unintentionally discriminate based on proxy information, AI applications are more susceptible to such discrimination due to their inherent structure [87]. Predictive AIs are designed to identify correlations between input data and target variables, but unlike traditional statistical models, they do not rely on human intuition about causal explanations. Instead, they use training data to identify characteristics that correlate with the target variable autonomously. This process disregards causation and leads AIs to inevitably seek out proxies for directly predictive characteristics when data on these characteristics is unavailable due to legal constraints. Merely denying AIs access to the most obvious proxies does not prevent this; it simply prompts them to rely on less intuitive proxies [87].

As a result, it becomes evident that proxy discrimination stemming from AI applications introduces significant limitations to the scope of anti-discrimination law. This occurs when an individual is discriminated against based on information other than those officially recognized as protected grounds. Instead of expanding the list of protected grounds, some scholars proposed as a solution to consider discrimination by association, where an individual associated with someone from a protected group could be covered [1,85]. For instance, if an AI application processes behavioral data to assess a job applicant's productivity and rejects the application of someone with a family history of depression and a personal history of sick leaves in prior jobs.

Given the timeliness of technological advancements and the contemporary understanding of certain concepts, however, the approach of expanding such a list of protected grounds should not be disregarded easily. For instance, emotion recognition systems, increasingly used in the workplace, aim to capture the inner states of a person [88]. Describing emotions is challenging, and attempts to use 'objective' technical methods, such as lie detectors or gender classifier systems, often lead to errors. If the assumption that physiological states directly impact user behaviors holds, severe consequences may arise [89]. Accordingly, there is a growing consideration among scholars about classifying emotion data as a new special category of personal data [89] and a new instance of privacy [90].

In a broader context, Sandra Wachter [91] emphasizes that discrimination in AI applications extends beyond traditional individual power imbalances and should encompass social groups currently lacking legal protection that share personal characteristics that can be either easily identifiable or not (human-comprehensible or incomprehensible characteristics). Examples include single parents, homeless individuals, and those with similar web histories or mouse movements. Addressing this limitation may necessitate legislative changes to redefine or expand the current list of protected grounds and protected target groups, ensuring AI applications adhere to specific requirements such as stability, transparency, empirical coherence, and ethical and normative acceptability [91].

Recognizing the emergence of new protected grounds and new affected groups and individuals in the context of anti-discrimination law could be instrumental in elevating the legal certainty with respect to what constitutes discrimination and what does not in an AI-dominated era. However, such efforts should also overcome another challenge of anti-discrimination law in addressing diversity biases in AI applications, namely its failure to account for intersectionality. Intersectional discrimination occurs when an individual experiences cumulative discrimination based on multiple personal characteristics interacting simultaneously [92]. At present, anti-discrimination law usually relies on single distinct protected grounds, and EU case law has traditionally denied the interaction of multiple systems of social disadvantage. Court rulings on gender and disability affecting maternity leave or age and sexual orientation in pension allocation have clearly illustrated this point [85]. With intersectional discrimination already prevalent in the analog world [23], deploying AI applications is expected to amplify and introduce new forms, given the datasets structured along intersecting axes of social inequality [85].

4.2. Data protection law

AI applications frequently involve the processing of personal data. This data can either contribute to forming the dataset used for training AI applications or be processed directly by the technology to fulfill specific tasks and objectives [93]. For recruitment and selection purposes, these AI applications are likely trained on personal data gathered from previous job vacancies, applications, and matches, which are then used to assess, categorize, score, or make decisions regarding the personal data of current job applicants [33,37]. As a result, data protection law could serve as an alternative or complementary framework to anti-discrimination law in defining and ensuring the fairness of AI applications for recruitment and selection. In the European Union, that framework is the General Data Protection Regulation (EU) 2016/679 (GDPR).

Within this context, Article 5(1)(a) of GDPR requires that personal data be processed 'fairly'. In doing so, the GDPR presupposes the subordinated position of the data subject to the data controller and incorporates the principle of fairness as a legal safeguard against this vulnerability [94]. Procedurally, this means that data processing is 'fair' if it aligns with the data subject's reasonable expectations, avoids unjustifiably negative impact, and allows the exercise of their rights [95]. However, fairness is just one element, with its effectiveness stemming from other GDPR measures. In this regard, a growing body of literature refers to many GDPR provisions influencing the design and development of fair AI applications in the labor market, particularly for selection and recruitment. This extends to other fundamental principles relating to the processing of personal data included in Article 5 GDPR, the performance of a data protection impact assessment (DPIA) under Article 35 GDPR, and the rights of the data subjects from Articles 12 to 22 GDPR.

Article 5 of the GDPR lays down all the key principles providing the basis for the protection of personal data [96], meaning that they could serve as the first cornerstone for ensuring fair data processing in the hiring context [79]. Besides fairness, these principles include lawfulness and transparency, thereby requiring compliance with legal requirements and providing information about the data processing to the job applicant. Purpose limitation directs the use of AI applications to specific, lawful recruitment goals, thereby preventing unintended data use. Data minimization ensures AI applications selectively process pertinent information, respecting candidate privacy. The accuracy principle emphasizes precise AI assessments to reduce biased decision-making. Storage limitation mandates retaining the personal data of job applicants only as necessary for recruitment. Integrity and confidentiality demand robust security measures in AI-driven recruitment to safeguard against unauthorized access. Accountability requires clear responsibilities for AI oversight, ensuring transparency and compliance with GDPR principles in hiring.

Article 35(1) GDPR mandates a DPIA when personal data processing poses a high risk to the rights and freedoms of individuals. The design and deployment of AI applications for recruitment and selection are likely to carry such risks due to their significant impact on job applicants' dignity, autonomy, and well-being and, more specifically, their social participation, economic circumstances, housing opportunities, family dynamics, as well as potential impacts on physical and mental health. The DPIA involves identifying and assessing risks to individuals' fundamental rights and freedoms and formulating a mitigation plan. Recent research suggests that conducting a DPIA can help organizations comply with the GDPR, demonstrate the fairness of their AI applications for recruitment and selection, and proactively identify and mitigate potential diversity biases [1,79]. However, one of the shortcomings of existing impact assessments is that they are bound to the company and only accessible as per request of, in this case, the DPA. By doing so, impact assessments lose the policy-relevant data-generating power they could have in showcasing best practices in given domains [97,98]. Although a good idea, the fact that these instruments are not connected to policymaking via a shared data repository, little is known about the practical realizations of the normative goals of regulatory instruments such as the GDPR.

Given that AI applications often autonomously evaluate and exclude job applicants without human intervention, it has been proposed that Articles 13–15 GDPR and Article 22 GDPR could play a pivotal role [1,4,38,99,100]. Articles 13 and 14 of the GDPR include the right to information, obligating data controllers to provide data subjects with "meaningful information about the logic involved, as well as the significance and envisaged consequences of such processing for [them]" [1,38]. Article 22 of the GDPR addresses the right of data subjects not to be subject to decisions "based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." Here the term 'decision' encompasses a broad scope, to the extent that the mere act of scoring is considered a decision (Schufa Holding, C-634/21). This implies that AI-driven hiring decisions are generally permissible, although they could be restricted if a job applicant invokes their right [99]. Alternatively, this provision is considered a general prohibition on automated decisions, except for limited exceptions [101,102]. This interpretation is supported by both the European Data Protection Board [103] and the ECJ case law (Schufa Holding, C-634/21). Article 22(2) GDPR provides exceptions that can legitimize automated decision-making in the recruitment and selection process, such as the necessity for entering or performing a contract and the individual's explicit consent. When implementing automated decision-making in line with these exceptions, it is crucial to ensure the explicit consent of the job applicant, which should be freely given, informed, specific, and unambiguous. This is particularly vital due to the inherent power imbalance in the relationships between the job applicant and the HR practitioner representing the employer [104]. In the labor market, where consent is marked by information asymmetry, decision-making authority, resource control, and economic dependence, a nuanced interplay of coercion and choice must be carefully considered [105]. Regardless, Article 22(3) of the GDPR requires data controllers to implement measures to safeguard data subjects' rights and freedoms, particularly the right to human intervention, expression of their viewpoint, and contesting the decision. In this regard, the ECJ has stressed that the data controller must actively inform the data subject about automated decision-making processes (Schufa Holding, C-634/21). The absence of strategies to identify and mitigate diversity biases from the controller side could violate this provision [79].

In cases where GDPR provisions are violated, national data protection authorities (DPAs) are responsible for monitoring and enforcing the regulation under Article 57(1)(a) GDPR. DPAs have a range of enforcement instruments at their disposal to promote fairness [79], such as requesting relevant information, accessing personal data, conducting audits according to Article 58(1) of the GDPR, and enforcing bias identification and minimization strategies through corrective powers under Article 58(2) of the GDPR. Additionally, substantial administrative fines can be imposed under Article 83 of the GDPR for breaches of these provisions [79].

However, the idea of relying on data protection law to address diversity biases in AI applications in the labor market has faced criticism and calls for caution. In particular, Frederik Zuiderveen-Borgesius [100] raises some critical points. In brief, compliance and enforcement deficits persist, and DPAs are often overburdened with limited sanctioning powers. Additionally, the scope of the GDPR is limited since it applies only to personal data and, for example, excludes predictive models lacking individual identification. While flexible, using open and abstract norms in data protection law can pose practical challenges. For example, the complexity of decision logic derived from extensive data analysis hinders the right to explain algorithmic decisions. At the same time, some tension arises from the imperative to adhere to Article 9 of the GDPR, which includes strict rules on processing special categories of data and the necessity to collect such information for effectively addressing and mitigating discrimination in algorithmic systems.

Briefly, the hiring process involves the collection and processing of personal data, such as racial or ethnic origins, which are so sensitive that they pose a risk of adversely affecting the fundamental rights of job applicants and carry a high risk of harm to them. Consequently, Article 9 GDPR includes a general prohibition on processing these special categories of data, with exceptions including explicit consent from the data subject (namely, the job applicant) and the fulfillment of obligations or exercise of specific rights by the controller (namely, the HR practitioner or employer) or the data subject (namely, the job applicant) in the field of employment [106]. At first sight, this provision could safeguard job applicants against diversity bias stemming from their personal characteristics, thereby ensuring the fairness of the recruitment and selection process. This is especially true when considering Article 9 of the GDPR in conjunction with the principles of data minimization and purpose limitation under Article 5 of the GDPR, which expects HR practitioners to collect only information relevant to the job application. Nonetheless, as previously reported, tensions may arise. On the one hand, the voluntary nature of consent and the necessity of data protection for employment purposes may be questioned due to the power imbalance between HR practitioners and job applicants [104], as applicants may fear rejection if they refuse consent or object to the necessity of data processing. Moreover, it has been increasingly argued that to assess whether an AI application unfairly discriminates against job applicants based on certain personal characteristics, the technology should process this information to identify and mitigate diversity bias [107].

4.3. The proposal for an AI Act

At the time of writing, EU policymakers were working on a proposal for a regulation laying down harmonized rules on artificial intelligence. In brief, the AI Act aims to establish a set of minimum rules to ensure the safe, lawful, and fair development, market introduction, and use of AI applications. At the same time, it seeks to uphold the principle of legal certainty, align with fundamental rights and safety standards, and facilitate the functioning of the single market. Accordingly, the proposal for an AI Act categorizes technology based on its design and associated risk levels, creating four distinct categories: unacceptable risks (Title II), high risks (Title III), limited risks (Title IV), and minimal risks (Title IX). As the risk level escalates, more stringent legal provisions are enacted [108].

The European Commission did not explicitly spell out any general principle regarding fairness. It took an instrumental and procedural approach to regulating AI applications, with the requirements for high-risk systems designed to prevent discrimination and respect fairness [109]. In its amendments adopted on 14 June 2023, instead, the European Parliament proposed the introduction of Article 4(a), which defines 'diversity, non-discrimination and fairness' as the development and use of AI applications "in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law." Besides, the European Parliament suggested rephrasing Recital 9, by including fairness among the values on which the Union is funded and should drive the technological design.

Given the novelty of the AI Act, its efficacy in ensuring fairness in the labor market remains somewhat uncertain, and there is limited research. For example, Antonio Aloisi [1] voices concerns about its preventive nature, suggesting it may lead to deregulation within the current sectoral and national legislative framework, which maintains a higher protection standard. Moreover, the Center for Democracy and Technology argues that the proposal for an AI Act currently lacks integration with existing safeguards provided by EU anti-discrimination law, potentially leaving gaps in protection when AI systems are used for recruitment. Additionally, the conformity assessment mechanism does not grant job applicants who experience discrimination any concrete rights to seek redress [110].

On a more general note, Recital No. 36 of the AI Act explicitly provides that:

"AI systems used in employment, workers management and access to self-employment, notably for the recruitment and selection of persons, for making decisions on promotion and termination and for task allocation, monitoring or evaluation of persons in work-related contractual relationships, should also be classified as high-risk, since those systems may appreciably impact future career prospects and livelihoods of these persons".

This means that any AI application for recruitment and selection must undergo a conformity assessment before its launch on the market. Compliance with essential requirements outlined in Chapter 2, Title III of the AI Act is crucial and encompasses data governance, technical documentation, record-keeping, transparency, user information provision, human oversight, robustness, accuracy, and security [111]. Overall, it appears that fairness is implicitly defined and ensured procedurally.

5. Fairness: Providing a bridge between disciplines and research communities

The previous section showed that anti-discrimination, data protection, and AI laws approach fairness piecemeal because they fail to offer a comprehensive and definitive definition and operationalization of fairness. This approach is wider than their legal framework governing fairness in AI applications for recruitment and selection. However, it extends to other applications, domains, and broader discussions, whether they originate from non-legal disciplines with context-sensitive viewpoints. Said otherwise, most literature on fairness tends to operate in isolated departments with limited cross-disciplinary communication [15]. For instance, the multi-faceted notion of fairness in computer sciences has been transposed in the high number of proposed mathematical and statistical definitions that nonetheless sometimes overlook the specific context of their possible applications or simply conflict with other metrics, like accuracy, transparency, and privacy [112,113]. Similarly, various philosophical theories offer distinct perspectives through which fairness can be examined without fully considering their potential operationalization and impact. *Inter alia,* these include formal, substantive, procedural, utilitarian, and distributive lenses [114–116].

In contrast to this piecemeal approach, some policymakers and scholars increasingly try to provide a bridge between different disciplines and research communities. Although, to our knowledge, there is little literature that does so in the context of AI applications for recruitment and selection purposes (*e.g.*, [77]), we refer to certain examples that might still be relevant, especially given the roadmap for future research and action unfolded in Section 6.

For example, the High-Level Expert Group on Artificial Intelligence (AI HLEG) regards fairness as an ethical principle later transposed into a key requirement, whether a technical or non-technical method.[1] As an ethical principle, fairness is seen through substantial and procedural lenses. Therefore, AI applications should guarantee equal and just distribution of benefits and costs while ensuring that each individual and group is free from unfair bias, discrimination, and stigmatization and could benefit from equal opportunities [117]. Besides, using AI applications should never lead to people being deceived or unjustifiably impaired in their freedom of choice. On the other hand, procedural fairness involves contesting and seeking effective redress against AI applications' decisions and those operating them. For this purpose, whoever is accountable for the decision must be identifiable, and the decision-making processes should be explicable [117]. Against this backdrop, the AI HLEG operationalizes fairness and intertwines it with diversity and non-discrimination. This implies that a fair AI application is expected to ensure equal access and treatment through an accessible system. inclusive design process engages with all those stakeholders who might directly or indirectly be affected through its life cycle [117].

In parallel with the AI HLEG, several scholars consider fairness a socio-technical problem and are engaged in delineating novel conceptualizations and operationalizations of fairness through multi-disciplinary frameworks. This is evident in the case of audits and impact assessments. In broad terms, both audits and impact assessments attempt to understand how the AI application plays out in practice, especially to evaluate its consequences on some legal, ethical, and social interests or values and foster accountability [5,112,118,119]. Similarly, Jia Qing Yap and Ernest Lim [120] suggest introducing an AI fairness reporting system, which requires revealing the utilization of all machine learning models, sharing information about the fairness metrics employed and the resulting compromises or choices made, providing

[1] The European Commission established the AI HLEG in June 2018 to support its roadmap for regulating AI. The AI HLEG first launched a consultation process and published the 'Ethics Guidelines for Trustworthy Artificial Intelligence'. For more information, *see* https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.

details about the methods employed to mitigate bias; and making datasets accessible for public examination. Otherwise, a growing body of literature refers to value- or legal-sensitive design and assumes that the technology design can significantly influence and nudge human behavior. Indeed, this perspective posits that ethical, moral, and legal norms can be effectively embedded into the software and hardware of AI applications through thoughtful and deliberate design practices [121–124]. In a similar way, some research develops and assesses fairness metrics against some legal rationales and objectives [125]. This is the case, for example, of Sandra Wachter et al. [126] who support bias-transforming metrics to promote substantive equality better, or Lisa Kutsoviti Koumeri et al. [127], who develop fairness constraints enabling a contextual approach to anti-discrimination case law in the EU. Foulds et al. [128], instead, propose an intersectional AI fairness criterion, which assumes that AI-driven outcomes are often influenced by social dynamics and encodes protection of each protected grounds of discrimination individually (*e.g.*, gender), as well as any of their subset (*e.g.*, gender and disability, gender and age, disability and age).

In bringing this section to a conclusion, we would like to highlight the reflections made by Hilde Weerts et al. [129], when discussing the complex relationship between law and computer science in defining and ensuring fairness. The authors assume that, rather than having some tree structure, "the law is dynamic, open-texture, and based on holistic reasonings" ([129], p. 814). This implies that its normative reasoning plays a fundamental role, and its anti-discrimination scope generally "fulfills a host of different social functions, ranging from the recognition of historical injustices and disadvantaged social groups, the (re)distribution of valuable goods and opportunities, the protection of dignity and autonomy, the accommodation of different lifestyles, and the facilitation of access to, and participation in, central social institutions such as the market, labor, education, healthcare, etc." ([129], p. 814). In this scenario, the authors continue, various normative objectives correspond to different conceptions of fairness, meaning that legal compliance cannot translate into a single threshold, fairness metric, or any other line of code and equation. Instead, starting thoughtfully and explicitly taking a normative stand on the final aim of legal and technical fairness interventions is necessary.

6. Future research and action: A cross-disciplinary and participatory approach

Although delving into a comprehensive normative framework is beyond the scope of this article, we will offer two practical recommendations and promising practices to guide further research and action on the design and use of AI applications for recruitment and selection, the ultimate aim also being to recognize, frame, and address the common asymmetry of power between job applicants and HR practitioners.

First, the pursuit of fairness of AI applications in the recruitment and selection process presents a multifaceted challenge, which encompasses broader and sometimes competing considerations from an ethical, legal, and technical angle. As it is hard to imagine a single research community possessing the all-encompassing knowledge required to navigate these complexities effectively, collaborative endeavors may allow for the pooling of resources, facilitating in-depth exploration of these intricacies in a cross-sector and cross-disciplinary way. For instance, the European Commission has funded several research projects to ensure fairness in the labor market through cross-disciplinary and cross-sector consortia involving experts from legal, social sciences, computer science, and other fields. For example, the BIAS project addresses diversity biases of AI applications for recruitment and selection by developing new and trustworthy technology for bias identification and mitigation and empowering the AI and HR communities through awareness raising and capacity building.[2] Similarly, the FINDHR project seeks to innovate in assessing discrimination risk, ensuring fair outcomes, and integrating human expertise in algorithmic hiring and other human recommendation systems, through effective procedures for software development, monitoring, and training [23].[3]

Second, cross-disciplinary and cross-sector synergies should rely on an inclusive and participatory approach. Different communities often harbor distinct stakeholder perspectives, which can vary based on the personal characteristics of individuals involved. Consequently, research teams should not only encompass a range of expertise but also embrace diversity in terms of gender, race, age, ability, sexual orientation, and other personal attributes that may influence perceptions of fairness. At the same time, proactive engagement with individuals who are not trained in research but belong to or represent whoever represents the target group or is potentially affected by the study. More precisely, this proactive engagement should encompass qualitative and quantitative research methods that are designed to facilitate participation, shared decision-making, and mutual learning and could, therefore, satisfy better the needs of scholars, research participants, and society at large [130,131]. In the said BIAS project, for example, the Consortium partners bring together cross-sector and multi-disciplinary expertise (*e.g.*, artificial intelligence, law, diversity studies, communication, and industrial commercialization). Additionally, they engage in extensive consultation and co-creation with a diverse pool of stakeholders, including HR practitioners, AI developers, policymakers, trade unions, civil-based society organizations, and academia. *Inter alia,* this participatory approach covers the performance of semi-structured interviews with AI developers and HR practitioners, the dissemination of a survey, some ethnographic fieldwork, and a series of co-creation workshops across Europe.

In conclusion, effectively addressing the multi-faceted challenge of ensuring fairness in AI applications for recruitment and selection needs a multi-disciplinary and participatory approach. While we acknowledge the difficulties associated with this, such as identifying appropriate representatives, ensuring balanced representation, and incorporating competing interests, the societal benefits far outweigh the challenges. This approach enables a more accurate understanding of technological innovation's ambitions and limitations in the labor market, facilitating the identification of suitable measures and interventions. Collaboration among diverse experts and stakeholders promotes capacity-building, knowledge exchange, and empowerment of people affected by AI applications in hiring, mainly when belonging to vulnerable and socially marginalized groups. This, in turn, enhances the social acceptability and sustainability of AI applications in the labor market.

7. Conclusions

The hiring process is increasingly driven by technological innovation, with AI applications more effectively identifying, attracting, screening, assessing, interviewing, or coordinating with job applicants. Nonetheless, the ambitious design of AI applications in the recruitment and selection process eventually carries certain flaws, including the risk of data protection violations and social discrimination. Given these caveats, we assumed that HR practitioners and employers started upholding fairness in each AI-driven hiring process, with this article examining this nebulous concept more in-depth.

Based on the cross-disciplinary literature addressing the divergent perceptions of fairness of all those stakeholders involved in the recruitment and selection process, it emerged that job applicants often recognize a common asymmetry of power in their interactions with HR

[2] For more information on the BIAS project: <https://www.biasproject.eu> accessed 18 October 2023

[3] For more information about the FINDHR project: <https://findhr.eu> accessed 18 October 2023

practitioners or potential employers. Job applicants tend to lean towards a procedural concept of fairness, wherein they expect recruitment and selection procedures to adhere to specific criteria, such as job relevance and consistency. Conversely, HR practitioners often prioritize finding the most suitable job applicant for the vacancy over ensuring fairness.

With this in mind, our attention turned toward the law as a possible venue for defining and guaranteeing the fairness of AI applications in the hiring process. In our analysis, we specifically delved into anti-discrimination, data protection, and AI laws governing AI uptake since they appeared to be the most pertinent regulations, given the impacts these technologies may have on workers. Certain consistent patterns emerged despite the inherent ambiguity surrounding the concept of fairness within the various regulations we examined. These included a shared objective of addressing power imbalances and the predominant procedural interpretation of fairness, often by fulfilling specified criteria. It also became apparent that current regulatory frameworks could benefit from revisiting, especially pondering the effectiveness and increased legal certainty of proxy discrimination and considering new protected discriminatory grounds and new protected social groups.

This apparent convergence of perceptions and interpretations might suggest a potential common understanding of fairness in AI applications for the recruitment and selection process. However, beneath the surface, this conclusion is somewhat illusory. Whereas there is agreement that fairness should address each social asymmetry of power and, as such, guarantee the full dignity, autonomy, and equality of individuals, the translation of this principle into procedural requirements introduces a complex interplay between the ideals and goals of the concept itself and their ulterior personal perception and interpretation. Briefly, the definitions of these requirements can vary depending on the stakeholders' expectations, aspirations, and concerns, the specific context of an application, and the interests at play. Therefore, the pursuit of procedural fairness runs a high risk of diverting attention and implementation from the fundamental normative stance that would ensure individuals' rights protection, which causes dissonances beyond the legitimacy and legality of practices in realizing fairness as a normative ideal.

Accordingly, we explored a growing body of literature that attempts to provide a bridge between different disciplines and research communities to define and operationalize fairness in AI applications. In this context, it was necessary to underscore the significance of embracing a multi-disciplinary and participatory approach in future research and action. This approach enriches our understanding and fosters collaborative efforts for a more legitimate, inclusive, and compelling exploration of fairness in AI applications for recruitment and selection.

Legislation

Amendments adopted by the European Parliament on the 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 (2023). P9_TA(2023) 0236

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 (2021). COM(2021) 206 final. 2021/0106(COD)

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016). L 119/1

Case-law

Centrum voor gelijkheid van kansen en voor racismebestrijding v. Firma Feryn NV (2008) Case no. C-54–07. European report of cases 2008 I-05,187.

Colin Wolf v. Stadt Frankfurt am Main (2010). Case no. C-229/08. European report of cases 2010 I-00,001.

Gorka Salaberria Sorondo v. Academia Vasca de Policía y Emergencias (2016). Case no. C-258/15. ECLI identifier: ECLI:EU:C:2016:873

Mario Vital Pérez v. Ayuntamiento de Oviedo (2014). Case no. C-416/13. ECLI identifier: ECLI:EU:C:2014:2371

Schufa Holding (Scoring) (2023). Case no. C-634/21. ECLI identifier: ECLI:EU:C:2023:957

Vera Egenberger v. Evangelisches Werk für Diakonie und Entwicklung e.V. (2018). Case no. C-414/16. ECLI identifier: ECLI:EU:C:2018:257

Declaration of competing interest

Carlotta Rigotti and Eduard Fosch-Villaronga - namely the authors of the manuscript 'Fairness, AI and Recruitment' - declare that they have no known competing interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgment

The authors are part of the Horizon Europe BIAS project, funded under the grant agreement no. 101070468. The information and views set out in this article are theirs and do not necessarily reflect the official opinion of the European Union.

References

[1] Aloisi A. Regulating algorithmic management at work in the European union: data protection, non-discrimination and collective rights. Int J Comp Labour Law and Ind Relat 2023;40(1):1–34.
[2] Ajunwa I. The "black box" at work. Big Data Soc 2020;7(2):205395172096618. https://doi.org/10.1177/2053951720938093.
[3] Black JS, van Esch P. AI-enabled recruiting: what is it and how should a manager use it? Bus Horiz 2020;63(2):215–26. https://doi.org/10.1016/j.bushor.2019.12.001.
[4] Ebert I, Wildhaber I, Adams-Prassl J. Big data in the workplace: privacy due diligence as a human rights-based approach to employee privacy protection. Big Data Soc 2021;8(1):205395172110130. https://doi.org/10.1177/20539517211013051.
[5] McKenzie R. Bots, bias and big data: artificial intelligence, algorithmic bias and disparate impact liability in hiring practices. Ark Law Rev 2018;71(2):529–70.
[6] Tambe P, Cappelli P, Yakubovich V. Artificial intelligence in human resources management: challenges and a path forward. Calif Manage Rev 2019;61(4): 15–42. https://doi.org/10.1177/0008125619867910.
[7] Li L, Lassiter T, Oh J, Lee MK. Algorithmic hiring in practice: recruiter and HR professional's perspectives on AI use in hiring. In: Proceedings of the 2021 AAAI/ACM conference on Ai, ethics, and society; 2021. p. 166–76. https://doi.org/10.1145/3461702.3462531.
[8] Chen Z. Collaboration among recruiters and artificial intelligence: removing human prejudices in employment. Cogn Technol Work 2023;25(1):135–49. https://doi.org/10.1007/s10111-022-00716-0.
[9] Drage E, Mackereth K. Does AI debias recruitment? Race, gender, and AI's "eradication of difference". Philos Technol 2022;35(4):89. https://doi.org/10.1007/s13347-022-00543-1.
[10] Köchling A, Wehner MC. Better explaining the benefits why AI? Analyzing the impact of explaining the benefits of AI-supported selection on applicant responses. Int J Sel Assess 2023;31(1):45–62. https://doi.org/10.1111/ijsa.12412.
[11] Varma A, Dawkins C, Chaudhuri K. Artificial intelligence and people management: a critical assessment through the ethical lens. Hum Resour Manag Rev 2023;33(1):100923. https://doi.org/10.1016/j.hrmr.2022.100923.
[12] Mujtaba DF, Mahapatra NR. Ethical considerations in AI-based recruitment. In: Proceedings of the 2019 IEEE international symposium on technology and society (ISTAS); 2019. p. 1–7. https://doi.org/10.1109/ISTAS48451.2019.8937920.
[13] Robles Carrillo M. Artificial intelligence: From ethics to law. Telecomm Policy 2020;44(6):101937. https://doi.org/10.1016/j.telpol.2020.101937.
[14] Carlsson V. Legal certainty in automated decision-making in welfare services. Public Policy Adm 2023:09520767231202334. https://doi.org/10.1177/09520767231202334.

[15] Mulligan DK, Kroll JA, Kohli N, Wong RY. This thing called fairness: disciplinary confusion realizing a value in technology. Proc ACM Hum Comput Interact 2019; 3(CSCW):1–36. https://doi.org/10.1145/3359221.
[16] Calleja C, Drukarch H, Fosch-Villaronga E. Harnessing robot experimentation to optimize the regulatory framing of emerging robot technologies. Data Policy 2022;4:e20. https://doi.org/10.1017/dap.2022.12.
[17] Autor DH. Wiring the labor market. J Econ Perspect 2001;15(1):25–40. https://doi.org/10.1257/jep.15.1.25.
[18] Cappelli P. Making the most of on-line recruiting. Harv Bus Law Rev 2001;79(3): 139–46.
[19] Bartram D. Internet recruitment and selection: kissing frogs to find princes. Int J Sel Assess 2000;8(4):261–74. https://doi.org/10.1111/1468-2389.00155.
[20] Hinton S, Schapper J. Jobs. com: recruiting on the net-a critical analysis of e-cruitment. Monash University; 2000. p. 1–18. Working Paper 56/00.
[21] Kaado B.. (2023). 12 LinkedIn Alternatives for Job Seekers. Business new daily. https://www.businessnewsdaily.com/8218-networking-sites-job-seekers.html.
[22] Sharone O. LinkedIn or LinkedOut? How social networking sites are reshaping the labor market. Emerging conceptions of work, management and the labor market, 30. Emerald Publishing Limited; 2017. p. 1–31. https://doi.org/10.1108/S0277-283320170000030001.
[23] Fabris, A., Baranowska, N., Dennis, M.J., Hacker, P., Saldivar, J., Borgesius, F.Z., & Biega, A.J. (2023). Fairness and Bias in Algorithmic Hiring. 10.48550/ARXIV.2309.13933.
[24] Albert ET. AI in talent acquisition: a review of AI-applications used in recruitment and selection. Strateg HR Rev 2019;18(5):215–21. https://doi.org/10.1108/SHR-04-2019-0024.
[25] Hemalatha A, Kumari PB, Nawaz N, Gajenderan V. Impact of artificial intelligence on recruitment and selection of information technology companies. In: Proceedings of the 2021 international conference on artificial intelligence and smart systems (ICAIS); 2021. p. 60–6. https://doi.org/10.1109/ICAIS50930.2021.9396036.
[26] Hussain R, Iqbal R. Elucidating the impact of cognitive and behavioral responses to web banner-ad frequency. J Mark Commun 2023:1–28. https://doi.org/10.1080/13527266.2023.2238206.
[27] Huang X, Yang F, Zheng J, Feng C, Zhang L. Personalized human resource management via HR analytics and artificial intelligence: theory and implications. Asia Pac Manag Rev 2023:S1029313223000295. https://doi.org/10.1016/j.apmrv.2023.04.004.
[28] Gonzalez MF, Liu W, Shirase L, Tomczak DL, Lobbe CE, Justenhoven R, Martin NR. Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. Comput Human Behav 2022;130:107179. https://doi.org/10.1016/j.chb.2022.107179.
[29] Naim I, Tanveer MdI, Gildea D, Hoque ME. Automated analysis and prediction of job interview performance. IEEE Trans Affect Comput 2018;9(2):191–204. https://doi.org/10.1109/TAFFC.2016.2614299.
[30] Drukarch H, Fosch-Villaronga E. The role and legal implications of autonomy in AI-driven boardrooms. B. Custers & E. Fosch-Villaronga (A c. Di). Law and artificial intelligence, 35. T.M.C: Asser Press; 2022. p. 345–64. https://doi.org/10.1007/978-94-6265-523-2_18.
[31] Raisch S, Krakowski S. Artificial intelligence and management: the automation–augmentation paradox. Acad Manag Rev 2021;46(1):192–210. https://doi.org/10.5465/amr.2018.0072.
[32] De Cremer D, Kasparov G. marzo 18). AI should augment human intelligence, not replace it. Harv Bus Rev 2021. https://hbr.org/2021/03/ai-should-augment-human-intelligence-not-replace-it.
[33] Acikgoz Y, Davison KH, Compagnone M, Laske M. Justice perceptions of artificial intelligence in selection. Int J Sel Assess 2020;28(4):399–416. https://doi.org/10.1111/ijsa.12306.
[34] Koivunen S, Olsson T, Olshannikova E, Lindberg A. Understanding decision-making in recruitment: opportunities and challenges for information technology. Proc ACM Hum Comput Interact 2019;3(GROUP):1–22. https://doi.org/10.1145/3361123.
[35] Derous E, De Fruyt F. Developments in recruitment and selection research: editorial. Int J Sel Assess 2016;24(1):1–3. https://doi.org/10.1111/ijsa.12123.
[36] Horodyski P. Recruiter's perception of artificial intelligence (AI)-based tools in recruitment. Comput Hum Behav Rep 2023;10:100298. https://doi.org/10.1016/j.chbr.2023.100298.
[37] Langer M, Baum K, König CJ, Hähne V, Oster D, Speith T. Spare me the details: how the type of information about automated interviews influences applicant reactions. Int J Sel Assess 2021;29(2):154–69. https://doi.org/10.1111/ijsa.12325.
[38] Sánchez-Monedero J, Dencik L, Edwards L. What does it mean to «solve» the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In: Proceedings of the 2020 conference on fairness, accountability, and transparency; 2020. p. 458–68. https://doi.org/10.1145/3351095.3372849.
[39] Sousa MJ, Wilks D. Sustainable skills for the world of work in the digital age: skills for the digital age. Syst Res Behav Sci 2018;35(4):399–405. https://doi.org/10.1002/sres.2540.
[40] Allal-Chérif O, Yela Aránega A, Castaño Sánchez R. Intelligent recruitment: how to identify, select, and retain talents from around the world using artificial intelligence. Technol Forecast Soc Change 2021;169:120822. https://doi.org/10.1016/j.techfore.2021.120822.
[41] Johnson RD, Stone DL, Lukaszewski KM. The benefits of eHRM and AI for talent acquisition. J Tour Futur 2021;7(1):40–52. https://doi.org/10.1108/JTF-02-2020-0013.
[42] Jarrahi MH. Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. Bus Horiz 2018;61(4):577–86. https://doi.org/10.1016/j.bushor.2018.03.007.
[43] Guenole N, Feinznig S. *The business case for AI in HR: Insights and tips on getting started*. IBM Smarter Workforce Institute, IBM Corporation; 2018. https://www.ibm.com/downloads/cas/A5YLEPBR.
[44] Cowgill, B. (2019). Bias and productivity in humans and machines. W.E. Upjohn Institute. 10.17848/wp19-309.
[45] Chamorro-Premuzic T, Akhtar R, Winsborough D, Sherman RA. The datafication of talent: how technology is advancing the science of human potential at work. Curr Opin Behav Sci 2017;18:13–6. https://doi.org/10.1016/j.cobeha.2017.04.007.
[46] Lee I, Shin YJ. Machine learning for enterprises: applications, algorithm selection, and challenges. Bus Horiz 2020;63(2):157–70. https://doi.org/10.1016/j.bushor.2019.10.005.
[47] Hunkenschroer AL, Kriebitz A. Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring. AI Ethics 2023;3(1):199–213. https://doi.org/10.1007/s43681-022-00166-4.
[48] Chowdhury S, Dey P, Joel-Edgar S, Bhattacharya S, Rodriguez-Espindola O, Abadie A, Truong L. Unlocking the value of artificial intelligence in human resource management through AI capability framework. Hum Resour Manag Rev 2023;33(1):100899. https://doi.org/10.1016/j.hrmr.2022.100899.
[49] Rampersad G. Robot will take your job: innovation for an era of artificial intelligence. J Bus Res 2020;116:68–74. https://doi.org/10.1016/j.jbusres.2020.05.019.
[50] Folger N, Brosi P, Stumpf-Wollersheim J, Welpe IM. Applicant reactions to digital selection methods: a signaling perspective on innovativeness and procedural justice. J Bus Psychol 2022;37(4):735–57. https://doi.org/10.1007/s10869-021-09770-3.
[51] Lee MK. Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. Big Data Soc 2018;5(1): 205395171875668. https://doi.org/10.1177/2053951718756684.
[52] Mirowska A, Mesnet L. Preferring the devil you know: potential applicant reactions to artificial intelligence evaluation of interviews. Hum Resour Manag J 2022;32(2):364–83. https://doi.org/10.1111/1748-8583.12393.
[53] Langer M, König CJ, Krause K. Examining digital interviews for personnel selection: applicant reactions and interviewer ratings. Int J Sel Assess 2017;25(4): 371–82. https://doi.org/10.1111/ijsa.12191.
[54] Butucescu A, Iliescu D. Patterns of change in fairness perceptions during the hiring process: a conceptual replication in a controlled context. Int J Sel Assess 2018;26(2–4):196–201. https://doi.org/10.1111/ijsa.12227.
[55] Konradt U, Garbers Y, Erdogan B, Bauer T. Patterns of change in fairness perceptions during the hiring process: patterns of change in fairness. Int J Sel Assess 2016;24(3):246–59. https://doi.org/10.1111/ijsa.12144.
[56] Konradt U, Oldeweme M, Krys S, Otte K. A meta-analysis of change in applicants' perceptions of fairness. Int J Sel Assess 2020;28(4):365–82. https://doi.org/10.1111/ijsa.12305.
[57] Gilliland SW. The perceived fairness of selection systems: an organizational justice perspective. Acad Manag Rev 1993;18(4):694. https://doi.org/10.2307/258595.
[58] Schinkel S, Van Vianen A, Van Dierendonck D. Selection fairness and outcomes: a field study of interactive effects on applicant reactions: selection fairness and outcomes. Int J Sel Assess 2013;21(1):22–31. https://doi.org/10.1111/ijsa.12014.
[59] Thorsteinson TJ, Ryan AM. The effect of selection ratio on perceptions of the fairness of a selection test battery. Int J Sel Assess 1997;5(3):159–68. https://doi.org/10.1111/1468-2389.00056.
[60] Zibarras LD, Patterson F. The role of job relatedness and self-efficacy in applicant perceptions of fairness in a high-stakes selection setting: selection fairness field study. Int J Sel Assess 2015;23(4):332–44. https://doi.org/10.1111/ijsa.12118.
[61] Alder GS, Gilbert J. Achieving ethics and fairness in hiring: going beyond the law. J Bus Ethics 2006;68(4):449–64. https://doi.org/10.1007/s10551-006-9039-z.
[62] Furnham A, Chamorro-Premuzic T. Consensual beliefs about the fairness and accuracy of selection methods at university: fairness and accuracy of selection methods at university. Int J Sel Assess 2010;18(4):417–24. https://doi.org/10.1111/j.1468-2389.2010.00523.x.
[63] Konradt U, Warszta T, Ellwart T. Fairness Perceptions in Web-based Selection: impact on applicants' pursuit intentions, recommendation intentions, and intentions to reapply: fairness in Web-based Selection. Int J Sel Assess 2013;21 (2):155–69. https://doi.org/10.1111/ijsa.12026.
[64] Truxillo DM, Steiner DD, Gilliland SW. The importance of organizational justice in personnel selection: defining when selection fairness really matters. Int J Sel Assess 2004;12(1–2):39–53. https://doi.org/10.1111/j.0965-075X.2004.00262.x.
[65] Gilliland SW, Groth M, Baker RC, Dew AE, Polly LM, Langdon JC. Improving applicants' reactions to rejection letters: an application of fairness theory. Pers Psychol 2001;54(3):669–703. https://doi.org/10.1111/j.1744-6570.2001.tb00227.x.
[66] Steiner DD, Gilliland SW. Procedural justice in personnel selection: international and cross-cultural perspectives. Int J Sel Assess 2001;9(1 & 2):124–37. https://doi.org/10.1111/1468-2389.00169.
[67] Truxillo DM, Bauer TN, Sanchez RJ. Multiple dimensions of procedural justice: longitudinal effects on selection system fairness and test-taking self-efficacy. Int J Sel Assess 2001;9(4):336–49. https://doi.org/10.1111/1468-2389.00185.

[68] Van Den Bos K, Vermunt R, Wilke HAM. Procedural and distributive justice: what is fair depends more on what comes first than on what comes next. J Pers Soc Psychol 1997;72(1):95–104. https://doi.org/10.1037/0022-3514.72.1.95.
[69] van Vianen, Taris R, Scholten E, Schinkel S. Perceived Fairness in Personnel Selection: Determinants and Outcomes in Different Stages of the Assessment Procedure. Int J Sel Assess 2004;12(1–2):149–59. https://doi.org/10.1111/j.0965-075X.2004.00270.x.
[70] Hilliard A, Guenole N, Leutner F. Robots are judging me: perceived fairness of algorithmic recruitment tools. Front Psychol 2022;13:940456. https://doi.org/10.3389/fpsyg.2022.940456.
[71] Koch-Bayram IF, Kaibel C, Biemann T, Triana MDC. Applicants' experiences with discrimination explain their reactions to algorithms in personnel selection. Int J Sel Assess 2023;31(2):252–66. https://doi.org/10.1111/ijsa.12417.
[72] Wesche JS, Sonderegger A. Repelled at first sight? Expectations and intentions of job-seekers reading about AI selection in job advertisements. Comput Hum Behav 2021;125:106931. https://doi.org/10.1016/j.chb.2021.106931.
[73] Roberson QM. The oxford handbook of diversity and work. Oxford University Press; 2013.
[74] Landon TE, Arvey RD. Ratings of test fairness by human resource professionals. Int J Sel Assess 2007;15(2):185–96. https://doi.org/10.1111/j.1468-2389.2007.00380.x.
[75] McAdams RH. The expressive powers of law: theories and limits. 1st ed. Harvard University Press; 2017. 1st printing.
[76] Nadler J. Expressive law, social norms, and social groups. Law Soc Inq 2017;42(01):60–75. https://doi.org/10.1111/lsi.12279.
[77] Hauer MP, Kevekordes J, Haeri MA. Legal perspective on possible fairness measures – a legal discussion using the example of hiring decisions. Comput Law Secur Rev 2021;42:105583. https://doi.org/10.1016/j.clsr.2021.105583.
[78] Adams-Prassl J, Binns R, Kelly-Lyth A. Directly discriminatory algorithms. Mod Law Rev 2023;86(1):144–75. https://doi.org/10.1111/1468-2230.12759.
[79] Hacker P. Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. Common Market Law Rev 2018;55(Issue 4):1143–85. https://doi.org/10.54648/COLA2018095.
[80] Baumgartner R, Arora P, Bath C, Burljaev D, Ciereszko K, Custers B, Ding J, Ernst W, Fosch-Villaronga E, Galanos V, Gremsl T, Hendl T, Kropp C, Lenk C, Martin P, Mbelu S, Morais Dos Santos Bruss S, Napiwodzka K, Nowak E, Williams R. Fair and equitable AI in biomedical research and healthcare: social science perspectives. Artif Intell Med 2023;144:102658. https://doi.org/10.1016/j.artmed.2023.102658.
[81] O'Neil C. Weapons of math destruction: how big data increases inequality and threatens democracy. 1st ed. Crown; 2016.
[82] Verhoef, T., & Fosch-Villaronga, E. (2023). Towards affective computing that works for everyone. 10.48550/ARXIV.2309.10780.
[83] Kim PT. Big data and artificial intelligence: new challenges for workplace equality. Univ Louiseville Law Rev 2019;57(2):313–28.
[84] Kleinberg J, Ludwig J, Mullainathan S, Sunstein CR. Discrimination in the Age of Algorithms. J Legal Anal 2018;10:113–74. https://doi.org/10.1093/jla/laz001.
[85] Xenidis R. Tuning EU equality law to algorithmic discrimination: three pathways to resilience. Maastrich J Eur Comp Law 2020;27(6):736–58. https://doi.org/10.1177/1023263×20982173.
[86] Bertrand M, Mullainathan S. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Am Econ Rev 2004;94(4):991–1013.
[87] Schwarcz D, Prince AER. Proxy Discrimination in the age of artifificial intelligence and big data. Iowa Law Rev 2020;105:1257–318.
[88] Boyd KL, Andalibi N. Automated emotion recognition in the workplace: how proposed technologies reveal potential futures of work. In: Proceedings of the ACM on human-computer interaction. 7; 2023. p. 1–37. https://doi.org/10.1145/3579528.
[89] Hauselmann, A., Sears, A.M., Zard, L., & Fosch-Villaronga, E. (2023). EU law and emotion data. 10.48550/ARXIV.2309.10776.
[90] Roemmich K, Schaub F, Andalibi N. Emotion AI at work: implications for workplace surveillance, emotional labor, and emotional privacy. In: Proceedings of the 2023 CHI conference on human factors in computing systems; 2023. p. 1–20. https://doi.org/10.1145/3544548.3580950.
[91] Wachter, S. (2022). The theory of artificial immutability: protecting algorithmic groups under anti-discrimination law. 10.48550/ARXIV.2205.01166.
[92] Crenshaw K. Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Univ Chic Leg Forum 1989;1(8):139–67.
[93] Sartor G. & Lagioia F.. (2020). The impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf.
[94] Malgieri G. The concept of fairness in the GDPR: a linguistic and contextual interpretation. In: Proceedings of the 2020 conference on fairness, accountability, and transparency; 2020. p. 154–66. https://doi.org/10.1145/3351095.3372868.
[95] Clifford D, Ausloos J. Data protection and the role of fairness. Yearb Eur Law 2018;37:130–87. https://doi.org/10.1093/yel/yey004.
[96] De Terwangne C. Article 5 principles relating to processing of personal data. C. De Terwangne. The eu general data protection regulation (GDPR). Oxford University Press; 2020. https://doi.org/10.1093/oso/9780198826491.003.0034.
[97] Fosch Villaronga E, Golia AJ. Robots, standards and the law: rivalries between private standards and public policymaking for robot governance. Comput Law Secur Rev 2019;35(2):129–44. https://doi.org/10.1016/j.clsr.2018.12.009.
[98] Fosch-Villaronga E, Heldeweg M. Regulation, I presume?" said the robot – Towards an iterative regulatory process for robot governance. Comput Law Secur Rev 2018;34(6):1258–77. https://doi.org/10.1016/j.clsr.2018.09.001.
[99] Parviainen H. Can algorithmic recruitment systems lawfully utilise automated decision-making in the EU? Eur Labour Law J 2022;13(2):225–48. https://doi.org/10.1177/20319525221093815.
[100] Zuiderveen Borgesius FJ. Strengthening legal protection against discrimination by algorithms and artificial intelligence. Int J Hum Rights 2020;24(10):1572–93. https://doi.org/10.1080/13642987.2020.1743976.
[101] Bygrave LA. Article 22 automated individual decision-making, including profiling. L. A. Bygrave. The eu general data protection regulation (GDPR). Oxford University Press; 2020. https://doi.org/10.1093/oso/9780198826491.003.0055.
[102] Tosoni L. The right to object to automated individual decisions: resolving the ambiguity of article 22(1) of the general data protection regulation. Int Data Priv Law 2021;11(2):145–62. https://doi.org/10.1093/idpl/ipaa024.
[103] Article 29 Data Protection Working Party. (2018). Guidelines on automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (17/EN WP251rev.01).
[104] Article 29 Data Protection Working Party. (2017). Opinion 2/2017 on data processing at work. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjy8cXh_ZiEAxWo9gIHHa-pBLQQFnoECBEQAQ&url=https%3A%2F%2Fec.europa.eu%2Fnewsroom%2Farticle29%2Fitems%2F610169&usg=AOvVaw3ejiko_xix0N74fO3S4OCG&opi=89978449.
[105] Niezna M, Davidov G. Consent in contracts of employment. Mod Law Rev 2023;86(5):1134–65. https://doi.org/10.1111/1468-2230.12802.
[106] Georgieva L, Kuner C. Article 9 processing of special categories of personal data. C. Kuner, L. A. Bygrave, C. Docksey, & L. Drechsler (A c. Di). The eu general data protection regulation (GDPR). Oxford University PressNew York; 2020. p. 365–84. https://doi.org/10.1093/oso/9780198826491.003.0038.
[107] Van Bekkum M, Zuiderveen Borgesius F. Using sensitive data to prevent discrimination by artificial intelligence: does the GDPR need a new exception? Comput Law Secur Rev 2023;48:105770. https://doi.org/10.1016/j.clsr.2022.105770.
[108] Veale M, Zuiderveen Borgesius F. Demystifying the draft EU artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach. Comput Law Rev Int 2021;22(4):97–112. https://doi.org/10.9785/cri-2021-220402.
[109] Pfeiffer J, Gutschow J, Haas C, Möslein F, Maspfuhl O, Borgers F, Alpsancar S. Algorithmic fairness in AI: an interdisciplinary view. Bus Inf Syst Eng 2023;65(2):209–22. https://doi.org/10.1007/s12599-023-00787-x.
[110] Ophélie S.. (2023). Discrimination in Hiring: the Case for Alignment of the EU AI Act with EU Equality Laws. https://cdt.org/insights/discrimination-in-hiring-the-case-for-alignment-of-the-eu-ai-act-with-eu-equality-laws/.
[111] Edwards L.. (2022). The EU AI Act: a summary of its significance and scope. Ada Lovelace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf.
[112] Calvi A, Kotzinos D. Enhancing AI fairness through impact assessment in the European Union: a legal and computer science perspective. In: Proceedings of the 2023 ACM conference on fairness, accountability, and transparency; 2023. p. 1229–45. https://doi.org/10.1145/3593013.3594076.
[113] Castelnovo A, Crupi R, Greco G, Regoli D, Penco IG, Cosentini AC. A clarification of the nuances in the fairness metrics landscape. Sci Rep 2022;12(1):4209. https://doi.org/10.1038/s41598-022-07939-1.
[114] Hooker B. Utilitarianism and fairness. B. Eggleston & D. E. Miller (A c. Di). The Cambridge companion to utilitarianism. 1a ed. Cambridge University Press; 2014. p. 280–302. https://doi.org/10.1017/CCO9781139096737.015.
[115] Rawls J, Kelly E. Justice as fairness: a restatement. Harvard University Press; 2001.
[116] Scanlon TM. Procedural fairness, 1. Oxford University Press; 2017. https://doi.org/10.1093/oso/9780198812692.003.0004.
[117] AI HLEG. (2019). Ethics Guidelines for Trustworthy AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html.
[118] Koshiyama A, Kazim E, Treleaven P, Rai P, Szpruch L, Pavey G, Ahamat G, Leutner F, Goebel R, Knight A, Adams J, Hitrova C, Barnett J, Nachev P, Barber D, Chamorro-Premuzic T, Klemmer K, Gregorovic M, Khan S, Lomas E. Towards algorithm auditing: a survey on managing legal, ethical and technological risks of AI, ML and associated algorithms. SSRN Electron J 2021. https://doi.org/10.2139/ssrn.3778998.
[119] Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: research methods for detecting discrimination on internet platforms. 64th annual meeting of the international communication association, seattle. http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20–%20Sandvig%20–%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf.
[120] Yap JQ, Lim E. A legal framework for artificial intelligence fairness reporting. Camb Law J 2022;81(3):610–44. https://doi.org/10.1017/S0008197322000460.
[121] Aizenberg E, van den Hoven J. Designing for human rights in AI. Big Data Soc 2020;7(2):205395172094956. https://doi.org/10.1177/2053951720949566.
[122] Hildebrandt M. Law for computer scientists and other folk. 1st ed. Oxford University Press; 2020.
[123] Simon J, Wong PH, Rieder G. Algorithmic bias and the value sensitive design approach. Internet Policy Rev 2020;9(4). https://doi.org/10.14763/2020.4.1534.
[124] van den Berg B, Leenes RE. Abort, retry, fail: scoping techno-regulation and other techno-effects. M. Hildebrandt & J. Gaakeer (A c. Di). Human law and computer

law: comparative perspectives. Netherlands: Springer; 2013. p. 67–87. https://doi.org/10.1007/978-94-007-6314-2_4.

[125] Schwartz R, Vassilev A, Greene K, Perine L, Burt A, Hall P. Towards a standard for identifying and managing bias in artificial intelligence (NIST sp 1270; p. nist sp 1270). National Institute of Standards and Technology (U.S.); 2022. https://doi.org/10.6028/NIST.SP.1270.

[126] Wachter S, Mittelstadt B, Russell C. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. Comput Law Secur Rev 2021; 41:105567. https://doi.org/10.1016/j.clsr.2021.105567.

[127] Koumeri LK, Legast M, Yousefi Y, Vanhoof K, Legay A, Schommer C. *Compatibility of Fairness Metrics with EU Non-Discrimination Laws: Demographic Parity & Conditional Demographic Disparity*. https://doi.org/10.48550/ARXIV.2306.08394.

[128] Foulds JR, Islam R, Keya KN, Pan S. An Intersectional Definition of Fairness. 2020 IEEE 36th International Conference on Data Engineering (ICDE) 2020;41: 1918–21. https://doi.org/10.1109/ICDE48307.2020.00203.

[129] Weerts H, Xenidis R, Tarissan F, Olsen HP, Pechenizkiy M. Algorithmic unfairness through the lens of eu non-discrimination law: or why the law is not a decision tree. In: Proceedings of the 2023 ACM conference on fairness, accountability, and transparency; 2023. p. 805–16. https://doi.org/10.1145/3593013.3594044.

[130] Cornish F, Breton N, Moreno-Tabarez U, Delgado J, Rua M, de-Graft Aikins A, Hodgetts D. Participatory action research. Nat Rev Methods Prim 2023;3(1):34. https://doi.org/10.1038/s43586-023-00214-1.

[131] Vaughn LM, Jacquez F. Participatory research methods – choice points in the research process. J Particip Res Methods 2020;(1):1. https://doi.org/10.35844/001c.13244.