Stanford – Vienna
Transatlantic Technology Law Forum

A joint initiative of
Stanford Law School and the University of Vienna School of Law

TTLF Working Papers

No. 91

Artificial Intelligence and the Future of Work: A Transatlantic Comparison on the Regulation of Algorithmic and Automated Decision-Making in the Workplace

Christine Carter

2022

TTLF Working Papers

Editors: Siegfried Fina, Mark Lemley, and Roland Vogl

About the TTLF Working Papers

TTLF's Working Paper Series presents original research on technology, and business-related law and policy issues of the European Union and the US. The objective of TTLF's Working Paper Series is to share "work in progress". The authors of the papers are solely responsible for the content of their contributions and may use the citation standards of their home country. The TTLF Working Papers can be found at http://ttlf.stanford.edu. Please also visit this website to learn more about TTLF's mission and activities.

If you should have any questions regarding the TTLF's Working Paper Series, please contact Vienna Law Professor Siegfried Fina, Stanford Law Professor Mark Lemley or Stanford LST Executive Director Roland Vogl at the

Stanford-Vienna Transatlantic Technology Law Forum
http://ttlf.stanford.edu

Stanford Law School
Crown Quadrangle
559 Nathan Abbott Way
Stanford, CA 94305-8610

University of Vienna School of Law
Department of Business Law
Schottenbastei 10-16
1010 Vienna, Austria

About the Author

Christine Carter holds both a BA (Double-First Class Honors) and a LLM (Distinction Honors) from the University of Cambridge, UK, where she has been awarded numerous academic prizes and studentships for her legal achievements. She has worked in the International Arbitration practice group of Arnold & Porter LLP in London where she assisted the practice group in various cases involving the representation of sovereign States and multi-national corporations in investor-state arbitrations and international commercial arbitrations. Her research interests lie in the areas of EU Law, Law & Economics, and Employment Law. Within these areas, she focuses on the regulatory, economic and technical implications of emerging technologies, particularly in relation to the regulation of Algorithmic Governance, Big Data, Digital Surveillance and Platform Economies.

General Note about the Content

Suggested Citation

This TTLF Working Paper should be cited as:
Christine Carter, Artificial Intelligence and the Future of Work: A Transatlantic Comparison on the Regulation of Algorithmic and Automated Decision-Making in the Workplace, Stanford-Vienna TTLF Working Paper No. 91, http://ttlf.stanford.edu.

Copyright

Abstract

The nature of work and status of workers has fundamentally changed over the past years in response to the rapid proliferation and introduction of Artificial Intelligence (AI) and Machine Learning (ML) technologies in the modern workplace, particularly in context of algorithmic and automated decision-making processes. These automated processes revolutionize the organization and management of human labor and are frequently used to make determinative decisions regarding the recruitment, performance and retention of workers. Against this background, concerns have arisen in relation to possible issues of algorithmic bias or discrimination, the maintenance of accountability and transparency, and the assurance of fairness and equity in the substantive decisions reached by these automated processes. This research paper investigates how the law responds to these concerns and regulates the use of AI and ML in automating decision-making processes within the modern workplace. The research paper adopts a comparative assessment of relevant labor protection laws in the United States and Europe, and focuses particularly on anti-discrimination and equality laws, data protection and privacy laws, as well as on the piecemeal legislative emergence of targeted AI regulations.

Keywords: Labor law, Artificial Intelligence, Machine Learning, Automated Decision-Making, Gig Economy

Table of Contents

Chapter 1: Introduction

Technological advancements in the workplace have historically resulted in a conflict between man and machine, which is visible as early as the 19th century when the automation of textile production led to a series of protests amongst English factory workers who plotted to destroy the machines during the proliferation of the Luddite movement[1]. Centuries on, the arrival of the Fourth Industrial Revolution was announced by Klaus Schwab, Executive Chairman, at the summit of world leaders in Davos[2]. Within this reality, work no longer resembles the post-industrialized idea of manual labor that is tethered to the physical workplace but represents a computerized and digitized recapitulation of Taylorism[3]. This can be seen in the introduction of Artificial Intelligence (AI) and Machine Learning Algorithms (ML) that have led to the emergence of algorithmic and automated decision-making in the workplace, whereby matters concerning the organization and management of the workforce are effectively shifted from humans to machines. The effect is to limit, and in some situations entirely replace, the role of human oversight in making such decisions with automated and algorithmic processes.

This has enormous significance and long-lasting impact on a worker's life, welfare and family[4]. For some, it will mean the difference between having and not having work. By way of a recent example, Unilever partnered with Pymetrics to create an online recruitment platform

---

[1] David H. Autor, 'Why Are There So Many Jobs? The History and Future of Workplace Automation'(2015) 29(3) 3-30 *The Journal of Economic Perspectives*, https://doi.org/10.1257/jep.29.3.3, at p. 3

[2] Klaus Schwab, *Shaping the Fourth Industrial Revolution* (Geneva World Economic Forum 2018), at pp. 11-12

[3] Moritz Altenried, 'The platform as factory: Crowdwork and the hidden labour behind artificial intelligence' (2020) 44(2) 145-158 *Capital & Class*, https://doi.org/10.1177/0309816819899410 at p.146; Dee Birnbaum and Mark Somers, 'Past as Prologue: Taylorism, the new scientific management and managing human capital' (2020) *International Journal of Organizational Analysis*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/IJOA-01-2022-3106

[4] European Trade Union Confederation 'Resolution on tackling new digital challenges to the world of labour, in particular crowdwork' (2017), https://www.etuc.org/en/document/etuc-resolution-tackling-new-digital-challenges-world-labour-particular-crowdwork at para 8; European Economic and Social Council 'Artificial intelligence-The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society' (2017), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016IE5369 at paras 1.10, 3.18-3.25

whereby job candidates complete a two-stage virtual suitability assessment that is operated by ML algorithms that first match the candidate's profile against those of the firm's current employees and then evaluate candidates in a video interview through a mixture of natural language processing and body language[5]. These methods, increase the likelihood of algorithmic bias and discrimination due to their reliance on iterative learning formulas, which may potentially even exclude groups of candidates that do not meet the encoded profile of a given algorithm[6]. Such in fact occurred through the controversial AI recruitment tool used by Amazon to hire technical staff, which turned out to discriminate against women because it had been trained on previous hiring decisions that were made over the past 10 years and happened at a time where the role of software development was a male dominated profession[7]. Obtaining legal redress in such situations is however often complicated by the complexity of legal review as a result of the evidential difficulties of gathering algorithmic data. AI automation tools are still in their technical infancy[8], and are often comprised of opaque algorithmic Black-boxes that make it impossible to ascertain the hidden logic behind the decision-making code[9]. These technical challenges have to be viewed against the backdrop of the proliferating casualization of work within the gig economy, where many workers are dependent on the mercy of the algorithms, particularly in a platform context where a bad rating by an algorithm will

[5] Bernard Marr 'The Amazing Ways How Unilever uses Artificial Intelligence to Recruit & Train Thousands of Employees' *Forbes* (New York, 14 December 2018) https://www.forbes.com/sites/bernardmarr/2018/12/14/the-amazing-ways-how-unilever-uses-artificial-intelligence-to-recruit-train-thousands-of-employees/?sh=6a113156274d at paras 5-7

[6] Joseph B. Fuller, Manjari Raman, Eva Sage-Gavin, Kristen Hines, et al 'Hidden Workers: Untapped Talent' (2021 Harvard Business School Project on Managing the Future of Work and Accenture https://www.hbs.edu/managing-the-future-of-work/Documents/research/hiddenworkers09032021.pdf at pp. 8-12

[7] Gina Neff, Maggie McGrath & Nayana Prakash, *AI @ Work* (2020) https://www.oii.ox.ac.uk/wp-content/uploads/2020/08/AI-at-Work-2020-Accessible-version.pdf at pp.9-10

[8] Michele Loi, 'People Analytics must benefit the people. An ethical analysis of data-driven algorithmic systems in human resources management' (2020) AlgorithmWatch, https://algorithmwatch.org/de/wp-content/uploads/2020/03/AlgorithmWatch_AutoHR_Study_Ethics_Loi_2020.pdf at p. 4

[9] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge: Harvard University Press, 2015), at pp.6-8

disconnect workers from their ability to obtain work and ensure income[10]. This raises significant accountability and transparency issues within the workplace. Specifically, should companies using AI to automate decisions be required to explain how, why, and what grounds a particular decision is reached? And in similar vein, does it make sense for the law to insist on a human-in-the-loop? These questions are intrinsic to the determination of whether the law is able to review an automated decision process as fair. They are also premised on whether the law is willing to recognize computational thinking as iterative and correlative processes and distinguish such from the intuitive and causative process of human thinking when considering questions of liability for algorithmic and automated decision-making.

The research paper will assess the regulatory responses of the United States and the European Union towards these emerging technologies to learn how each jurisdiction tackles these technological changes and understand what can be learnt from their comparative approaches. This will explain how AI impacts both the future of work and the capacity of law, as an organizational framework, to institute the industrial relationship between labor and capital, as society transitions from a post-industrialist to a new-digital era.

The following structure will be adopted: *Chapter 2* exposes the background and regulatory significance of AI in automated decision-making in the workplace. *Chapter 3* compares the regulation of AI in automated decision-making under EU and US anti-discrimination laws. *Chapter 4* then compares how this is achieved under EU and US data protection and privacy laws, and *Chapter 5* examines the extent to which both jurisdictions have introduced targeted AI regulations. *Chapter 6* concludes.

[10] Foundation for European Progressive Studies (FEPS) 'Work in the European Gig-Economy' (2017) https://uhra.herts.ac.uk/bitstream/handle/2299/19922/Huws_U._Spencer_N.H._Syrdal_D.S._Holt_K._2017_.pdf at p.13

Chapter 2: The Algorithmic and Automated Workplace

AI performs a key role in the modern workplace and interacts with all levels of decision-making in regards to the organization and management of labor. Algorithmic and automated decision-making is pervasively used within a variety of contexts in the workplace, especially in recruitment, retention, performance review decisions[11], as well as in the allocation of work and determination of compensation levels[12]. The Committee of Experts and Internet Intermediaries observe how such uses of *"mathematic or computational constructs do not by themselves have adverse human rights impacts but their implementation and application to human interaction does"*[13]. Indeed, the act of limiting, or even replacing, the role of human agency in decision-making raises a distinct set of regulatory challenges in relation to issues of bias, discrimination, transparency and accountability . These issues are crucially distinct to the human-to-human context of labor organization and management and may therefore be harder to detect due to their novelty in presentation and more unusual forms of manifestation both within and beyond, the workplace context.

I. Organizational Uses of AI Decision-Making

I.A. Recruitment Decisions

AI is often used as a tool to target particular job-advertisement to specific candidates based on a technique called candidate-profiling whereby the algorithm effectively acts as a head hunter. This occurs at the outreach stage of the recruitment process where AI identifies

[11] Trade Union Congress 'Technology Managing People: The Worker Experience Report' (2020), https://www.tuc.org.uk/research-analysis/reports/technology-managing-people-worker-experience at pp. 17-36

[12] Centre for Data Ethics and Innovation 'Review into bias in algorithmic decision-making' (2020), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957259/Review_into_bias_in_algorithmic_decision-making.pdf at pp.39-48

[13] Committee of Experts on Internet Intermediaries 'Algorithms and human rights - Study on the human rights dimensions of automated data processing techniques and possible regulatory implications' (2018) https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html at p.8

prospective candidates externally from professional job search and networking sites such as LinkedIn or alternatively from internally automated tracking systems[14]. AI is also frequently used at the screening stage of the recruitment process to shortlist candidates and filter out unsuccessful candidates to save time and costs that would otherwise be incurred if HR were to sieve through all candidates[15]. In an interview with the Guardian Newspaper, Ian Siegel, CEO of ZipRecruiter, stated that about three-quarters of resumes submitted for jobs in the US are now being read and reviewed by AI[16]. These technologies rank candidates through algorithms that make predictions on the candidate's likely future job performance[17]. In the later stages of the process, audio and visual software is often used in interviews to assess the candidates' performance and their suitability for a particular position[18]. Similar can be seen in the use of AI stimulations, in which job applicants are required to complete a gamified assessment that evaluates the candidate's behavior and extrapolates this data into a success profile[19].

I.B. Review Decisions

The use of AI in performance reviews is very multifaceted, with firms using algorithms to monitor the location, productivity, task-distribution and working hours among workers[20].

---

[14] Patrick van Esch, Stewart Black, 'Factors that influence new generation candidates to engage with and complete digital, AI-enabled recruiting' (2019) 62(6) 729-739 *Business Horizons,* https://doi.org/10.1016/j.bushor.2019.07.004

[15] Carmen Fernández, Alberto Fernández 'AI and recruiting software: Ethical and legal implications' (2020) 11(1) 199-216 *Paladyn: Journal of Behavioral Robotics* https://doi.org/10.1515/pjbr-2020-0030 at p.204; Marisa Vasconcelos, Carlos Cardonha, Bernardo Gonçalves, B 'Modeling epistemological principles for bias mitigation in AI systems: An illustration in hiring decisions' (2018) *AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA, February*, 323–329, New York: ACM. https://doi.org/10.1145/3278721.3278751.

[16] Hilke Schellmann 'Finding it hard to get a new job? Robot recruiters may be to blame' *Guardian News* (London, 11 May 2022) https://www.theguardian.com/us-news/2022/may/11/artitifical-intelligence-job-applications-screen-robot-recruiters at para 7

[17] Miranda Bogen 'All the ways hiring algorithms can introduce bias' (2019) *Harvard Business Review* https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias at para 14

[18] Alina Köchling et al. 'Highly accurate, but still discriminatory: A fairness evaluation of algorithmic video analysis in the recruitment context' (2021) *Business Information Systems Engineering* 63(1) 39-54 https://doi.org/10.1007/s12599-020-00673-w at p.41

[19] Prasanna Tambe, Peter Cappelli et al 'Artificial intelligence in human resources management: Challenges and a path forward' (2019) *California Management Review,* 61(4) https://doi.org/10.1177/0008125619867910 at pp.16-17

[20] Valerio De Stefano 'Negotiating the Algorithm: Automation, Artificial Intelligence and Labor Protection' (2019) 41(1) *Comparative Labor Law & Policy Journal,* http://dx.doi.org/10.2139/ssrn.3178233 at pp.10-17

Many algorithms also rate workers and deliver targeted recommendations to specific workers on how they can improve performance[21]. This particular function is often termed 'algorithmic management'[22]. The idea behind this term is that it denotes the managerial and organizational function of the algorithm to explain its increasing dominance in crowdsourcing platforms that can be seen in companies such as TaskRabbit, Uber, Lyft and Deliveroo, in which the algorithmic platform acts as an online intermediary that allocates casual work to workers who aren't necessarily employed by these platforms. In other contexts, algorithmic management can also be seen in its operation as a performance-feedback provider. Enaible offers an example of performance-feedback AI programs. The AI platform monitors the performance of remote workers by giving them a productivity score and then identifying ways for them to increase this scoring through an AI feedback program[23]. Similar can be seen in the AI program used by the insurance company MetLife that uses advanced data analytics to track worker's conversations with customers and then uses this information to provide personalized recommendations to workers on how they can improve upon their interactions with customers and get better results at work[24]. The data derived from performance monitoring AI is not only limited to providing customized feedback but can be used as input in decisions that would ordinarily be delegated to line-management, such as decisions on work allocation and pay.

[21] Katherine C Kellogg et al. 'Algorithms at work: the new contested terrain of control' (2020) 14(1) 366-41- *Academy of Management Annals* https://doi.org/10.5465/annals.2018.0174 at p.372
[22] Jeremias Prassl, *Humans as a Service: The Promise and Perils of Work in the Gig Economy* (OUP 2018), at pp.5-6
[23] Will Douglas Heaven 'This startup is using AI to give workers a "productivity score"' (2020) *MIT Technology Review* https://www.technologyreview.com/2020/06/04/1002671/startup-ai-workers-productivity-score-bias-machine-learning-business-covid/ at paras 5-7
[24] Kevin Roose, 'A machine may not take your job, but one could become your boss' (New York, 2021) *The New York Times* https://www.nytimes.com/2019/06/23/technology/artificial-intelligence-ai-workplace.html; ACAS 'My boss the algorithm: an ethical look at algorithms in the workplace' (2020) https://www.ipa-involve.com/Handlers/Download.ashx?IDMF=7129b512-2368-459a-898d-6d2b3457a039 at pp.13-16

I.C. Retention Decisions

The ability of AI to assess workers on a targeted set of criteria naturally lends itself to its utility as a decision-maker in situations involving the dismissals, redundancies and transfers of undertakings of workers. These algorithms not only monitor the workers but also are able to provide rankings based on attained, or even future performance[25]. This particular use of automated decision-making can be seen in the dismissal of Amazon Drivers following the surveillance, monitoring and data-analysis service of the new Amazon Flex App[26]. The App used a series of algorithms that measured the drivers' delivery patterns, their adherence to the prescribed route, and their punctuality. Based on the results, the drivers are then either given more work or are no longer engaged. These types of Algorithms are not only considered more efficient in reaching the decision on who should and who shouldn't get fired but also in executing the decision. To give an example; in 2020, the Russian Software company Xsolla instantaneously dismissed 150 out of its 450 employees, based solely on the results reached by an AI an automated decision on the workforces' engagement on productivity standards[27].

II. Regulatory Risks of AI Decision-Making

[25] European Commission 'Algorithmic Management Consequences for Work Organisation and Working Conditions' (2021) *JRC Working Papers Series on Labour, Education and Technology 2021/07* https://joint-research-centre.ec.europa.eu/publications/algorithmic-management-consequences-work-organisation-and-working-conditions_en at pp.4-8

[26] Spencer Soper 'Fired by Bot at Amazon: 'It's you against the Machine'' (28 June 2021) *Bloomberg News* https://www.bloomberg.com/news/features/2021-06-28/fired-by-bot-amazon-turns-to-machine-managers-and-workers-are-losing-out

[27] Miquel Echarri, 'One second, 150 dismissals: Inside the algorithms that decide who should lose their job' (14 October 2021) *El País* https://english.elpais.com/usa/2021-10-14/one-second-150-dismissals-inside-the-algorithms-that-decide-who-should-lose-their-job.html

II.A. Bias

AI bias arises where algorithms exhibit a particular prejudice or an adverse inclination towards particular individuals[28]. AI bias comes within three forms.

First, there is *Data Set Bias*[29] where the training data set does not adequately represent a diverse or representative user-base. For instance, if an algorithm is constructed on the ethnography of the existing employee population at a given company that is not representative or inclusive of particular social groups, then there is a danger that the algorithm will reproduce this underrepresentation when recruiting future candidates[30]. This often emerges from sampling errors such as the collection of data from a skewed sample, over and under-sampling, limited feature choices within the sample, proxies and redundant encodings and human bias behind the sampling collection.

Second, there may be *Model Design Bias*. Protected characteristics, such as race, gender, age, ethnicity or other, may also be encoded into the algorithmic model as target variables. Even where the variables are audited to remove protected characteristics, bias can alternatively be introduced via the design models in situations where the model identifies unanticipated proxies that are associative of protected characteristics. For instance, a model may analyze the zip codes of job candidates to the workplace to ensure that they can work optimal hours if so required. The zip codes of the candidates may be a proxy for discrimination since different neighborhoods have different population backgrounds.

---

[28] Lynette Yarger, Faye Cobb Payton and Bikalpa Neupane 'Algorithmic equity in the hiring of underrepresented IT job candidates' (2019) *Online Information Review,* 44(2) 383-395. https://doi.org/10.1108/OIR-10-2018-0334

[29] Manish Raghavan et al 'Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices' (2019) 469-481 *Proceedings on the 2020 Conference on Fairness, Accountability and Transparency* https://doi.org/10.1145/3351095.3372828

[30] Ketki V. Deshpande, Shimei Pan, and James R. Foulds. 'Mitigating Demographic Bias in AI-based Resume Filtering' (2020) (Adjunct Publication of the 28th ACM Conference on User Modelling, Adaptation and Personalization "UMAP '20 Adjunct" Association for Computing Machinery), New York, NY, USA, at pp. 268–275. https://doi.org/10.1145/3386392.3399569; European Disability Forum, *Plug and Pray? A disability perspective on artificial intelligence, automated decision-making and emerging technologies* (2018), Plug and pray? A disability perspective on artificial intelligence, automated decision-making and emerging technologies" at pp.26-27

Third, there may be *ML Self-Training Bias*[31], where the algorithm replicates, and potentially even amplifies the bias in the training data when making future predictions and selections. In this sense, bias can be either trained into the algorithm or the algorithm can alternatively learn from a bias sample. Such often occurs through the use of ML in facial recognition software where the algorithms perform less well when analyzing ethnic minorities when compared to white users because of the underrepresentation of this social group within its data-set, which is then fed into the training of the algorithm[32].

II.B. Discrimination

The use of AI in automating decisions risks systematizing, and potentially even multiplying, the input of human bias into the output of the processing system, which will lead to discriminatory results[33]. Crucially, however, and different to a human-to-human context of discrimination, discrimination in the AI context has a very different form[34]. AI discrimination is often dynamic in the sense that algorithms continuously develop new forms of categorizing individuals on the basis of new correlations and training set predictions and outcomes[35]. This

---

[31] Yarger et al, n 28 at p.385

[32] Michael Gentzel 'Biased Face Recognition Technology Used by Government: A Problem for Liberal Democracy' (2021) 34(4) 1639–1663 *Philosophy & Technology.* https://doi.org/10.1007/s13347-021-00478-z; Joy Buolamwini, and Timnit Gebru, 'Gender shades: Intersectional accuracy disparities in commercial gender classification' (2018) 81(1) 77-91 *Proceedings of Machine Learning Research at the Conference on Fairness, Accountability, and Transparency* https://proceedings.mlr.press/v81/buolamwini18a.html; Drew Harwell 'Contract lawyers face a growing invasion of surveillance programs that monitor their work' *The Washington Post* (Washington, 2021) https://www.washingtonpost.com/technology/2021/11/11/lawyer-facial-recognition-monitoring/ paras 17-20

[33] Solon Barocas and Andrew D. Selbst 'Big Data's Disparate Impact' (2016) 104(1) 671-723 *California Law Review,* http://dx.doi.org/10.2139/ssrn.2477899 at p. 677 and p.679

[34] Raphaële Xenidis and Linda Senden, 'EU Non-Discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination' in Ulf Bernitz and others (eds), *General Principles of EU law and the EU Digital Order* (Kluwer Law International 2020) at pp.151-182 https://ssrn.com/abstract=3529524

[35] Raphaële Xenidis 'Tuning EU equality law to algorithmic discrimination: Three pathways to resilience' (2020) Vol. 27(6) 736–758 *Maastricht Journal of European and Comparative Law* https://doi.org/10.1177/1023263X20982173 at p. 738; Anna Lauren Hoffmann 'Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse' (2019) 22(7) 900-915 *Information, Communication & Society*, 10.1080/1369118X.2019.1573912; See further Second Amended Class and Collective Action Complaint and Demand for Jury Trial Bradley v. T-Mobile US, Inc., No. 5:17-cv-07232 (N.D. Cal. Aug. 20, 2018); Pauline Kim 'Big Data and Artificial Intelligence: New Challenges for Workplace Equality' (2019) 57 *University of Louisville Law Review* (Forthcoming)

poses four new challenges for existing anti-discrimination and equality laws in terms of the conceptualization and treatment of discriminatory algorithms.

First, *AI discrimination is often intersectional* due to the multifactorial disposition of different variables. This is frequently caused by AI processes such as redundant encoding[36], and feedback loops and effects[37]. These processes have the effect that discrimination may not only impact one individual but entire classes of persons or potentially even multiple classes.

Second *AI discrimination is often proxy-based* and therefore associative rather than constitutive of particular features[38], which is exemplified in the aforementioned example of selecting employees on the basis of their home address[39]. In that situation, the zip-code would correlate with a protected class and therefore discriminate applicants from that class through association with the particular area.

Third, *AI discrimination is often non-causal*, due to the non-deterministic and probabilistic operation of algorithms that identify correlation and association between data variables instead of causation. This is due to the unintuitive and regression based data analysis of algorithms[40]. Even where causal methods are used to operate algorithms, they can often lead to a data apophenia[41] or result in what computer scientists call Simpson's paradox where trends observed in data in fact reverse when more data is accumulated, which in turn then leads to an unjustified outcome with potentially discriminatory results.

---

[36] Cynthia Dwork et al. 'Fairness through awareness' (2012) *Proceedings of the 3rd innovations in theoretical computer science conference,* pp.214-226 https://doi.org/10.1145/2090236.2090255

[37] Pauline T. Kim 'Data-Driven Discrimination at Work' (2017) 58(3) 857-936 *William & Mary Law Review* https://scholarship.law.wm.edu/wmlr/vol58/iss3/4; at p.866, p. 882, pp. 895-896 Anna Lauren Hoffmann 'Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse' (2019) 22(7) 900-915 *Information, Communication & Society*, 10.1080/1369118X.2019.1573912

[38] Anya Prince and Daniel Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2019) 105(1) 1257-1266 *Iowa Law Review* at p.1276

[39] Xenidis and Senden, supra n 34 at p. 155

[40] Sandra Wachter and Brent Mittelstadt and Charles Russel 'Why fairness cannot be automated : bridging the gap between EU non-discrimination law and AI' (2020) 41(2021): 105567 *Computer Law & Security Review* https://ssrn.com/abstract=3547922 at pp.10-13

[41] Kate Crowford and Danah Boyd 'Critical questions for big data' (2012) 15(5) 662-679 *Information Communication & Society* https://doi.org/10.1080/1369118X.2012.678878 at p. 668

Fourth and consequently, *AI discrimination is often novel,* which will potentially disrupt established patterns of discrimination and create new ones, which makes the detection of evidence that AI discrimination has occurred extremely difficult[42].

II.C. Transparency

Issues of proof are also often accompanied with a significant lack of awareness by workers that AI is even being used in the first place to automate workplace decisions, which perpetuates the often already precarious power dynamics within the workplace context even further[43]. This creates what Pasquale calls the so-called Black-Box[44] phenomenon because the automation of workplace decision-making by these algorithms is subject to a procedure that is not only often unknown and unexplainable to the workers and the employer, but often to the maker of the algorithm themselves, which creates the common perception that AI decisions are therefore 'untrustworthy'[45].Gaudio gives three reasons why the use of AI decision-making in the workplace is often opaque[46].

The first reason is *Legal Opacity*. AI is often regulated and owned by corporate and trade secrecy laws, intellectual property laws, and in particular the laws relevant to contractual

[42] Matthias Leese 'The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union' (2014) 45(1) 494-511 *Security Dialogue* https://www.qub.ac.uk/Research/GRI/mitchell-institute/FileStore/Filetoupload,756547,en.pdf at pp. 504-505; Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) *Big Data & Society* 10.1177/2053951715622512 at pp.1-12; Christopher Kuner et al. 'Machine Learning with Personal Data: Is Data Protection Law Smart Enough to Meet the Challenge?' (2017) 7 *IDPL* at p.1; Jennifer Cobbe & Jatinder Singh 'Regulating Recommending: Motivations, Considerations, and Principles' (2019) 10(3) *European Journal of Law & Technology*. https://papers.ssrn.com/ab- stract=3371830
[43] Alex Rosenblat and Luke Stark 'Algorithmic labor and Information Asymmetries: A Case Study of Uber's Drivers' (2016) 10(2016), 3758-378410 *International Journal of Communication* https://ijoc.org/index.php/ijoc/article/view/4892/1739 at p. 3759
[44] Pasquale, n 9 at pp. 6-8
[45] European Commission 'Ethics guidelines for trustworthy AI' (2019) https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. at p.13
[46] Giovanni Gaudio 'Algorithmic Bosses Can't Lie! How to Foster Transparency and Limit Abuses of the New Algorithmic Managers' (2021). *Comparative Labor Law & Policy Journal (*Forthcoming), Bocconi Legal Studies Research Paper No. 3927954 https://ssrn.com/abstract=3927954

and business confidentiality[47]. Although these laws protect algorithms as valuable economic commodities, they inhibit workers from understanding what rights they have over their data and its usage, which prevents them from making informed inquiries.

The second reason is *Code-Language Opacity*. Specialized skills, training and a vast pool of resources are required for humans to visualize and interpret large volumes of data and code [48]. Especially in the age of Big-Data, algorithms have to compute billions and billions data entities, which has led to the perpetuation of highly complex codes[49]. It is also noteworthy that code-language complexity is not always a concession to the sheer complexity of the data that algorithms need to process but sometimes even intentional. An example this can be seen in the Google PageRank algorithm, which is deliberately opaque, to prevent users from gaming the system and competitors from stealing its innovative potential[50].

The third reason is *ML Opacity*, which emerges between the trade-off between accuracy and interpretability when selecting suitable algorithms for data predictions. Rule based algorithms, such as linear regressions and decision trees don't create these opacities and are therefore less problematic. In contrast, ML algorithms are more accurate than rule based algorithms because of their non-linear and non-monotonic functions but are therefore much harder to explain or reverse-engineer since there is no certainty how much importance the algorithm places on each feature of the model prediction or its interaction with others. Similar problems also arise with neural networks and gradient boosting models.

---

[47] Afzana Anwer 'How SMEs can use IP to secure success in the new data-fuelled AI paradigm' *IAM News* (New York, 17 February 2021) https://www.iam-media.com/article/ip-opportunities-and-challenges-smes-in-the-new-data-fuelled-ai-paradigm

[48] Prasanna Tambe and Peter Cappelli, n 19 above at p.12

[49] Burrell, n 42 at pp.1-12

[50] John Naughton 'Good Luck in Making Google Reveal its Algorithm' *Guardian News* (London: November 2016) https://www.theguardian.com/commentisfree/2016/nov/13/good-luck-in-making-google-reveal-its-algorithm at para 2 specifically ("*…algorithms must be made more transparent, so that one can inform oneself as an interested citizen about questions like, 'What influences my behaviour on the internet and that of others?' Algorithms, when they are not transparent, can lead to a distortion of our perception; they can shrink our expanse of information.*")

These issues not only make algorithmic and automated decision-making in the workplace context difficult to understand but potentially impossible to challenge where an unfavorable result has been reached.

II.D. Accountability

Accountability touches upon the need for the automated decision to be responsible and explainable[51].

In relation to the issue of responsibility, it is crucial to consider the role of human oversight in automated decision-making and whether there is a need to have a human-in-the-loop who can act as an assurance system for the automated decision. This is in part due to concerns of 'agency laundering' where algorithms effectively distance the user from morally suspect actions, regardless of their intentions. The omission of liability contributes to a de facto 'computer said so' defense for the human behind the algorithm[52]. An example of such can be seen in the Facebook ProPublica Ad Scandal where Facebook's algorithm created categories that targeted ads to Anti-Semitic groups and the company sought to defend the creation of such on the basis that these were autonomously created by algorithms that had canvassed responses of Facebook users to specific target fields[53]. It is also in part due to the human propensity to over-rely on the perceived objectivity of automated and algorithmic decision-making[54]. Relatedly, this also touches on the issue of 'explainability'.

---

[51] Joana Hois, Dimitra Theofanou-Fuelbier and Alischa Janine Junk, 'How to Achieve Explainability and Transparency in Human AI Interaction'. In: Stephanidis, C. (eds) HCI International 2019 - Posters. HCII 2019. Communications in Computer and Information Science, 1033(1). Springer https://doi.org/10.1007/978-3-030-23528-4_25 at pp.177-183.

[52] *See for general discussion* Tero Karppi 'The Computer Said So: On the Ethics, Effectiveness and Cultural Techniques of Predictive Policing' (2018) 1-8 *Social Media and Society* 4(2) https://doi.org/10.1177/2056305118768296 at p. 2

[53] Alan Rubel 'Agency Laundering and Information Technologies' (2019) 22(1) 1017–1041 *Ethical Theory Moral Pract* https://doi.org/10.1007/s10677-019-10030-w at p.1018

[54] European Commission 'Proposal for a Council Directive concerning the protection of individuals in relation to the processing of personal data' (1990) [COM(92) 422 final – SYN 287], Explanatory memorandum, at para 26

For automated decision-making to be accountable, it must be possible to ask about and ascertain information relating to questions such as ‘what is the basis on which the algorithm is making particular choices’, ‘what data and variables is it using in order to conduct its data analysis’, ‘how does it process this data’ and ‘is there accountability for the decision-made’? Most job candidates are unaware of the use of AI hiring algorithms in processing their data, including information from their applications, CV, social-media sites and video applications, and had therefore never given any informed consent to the use of AI decision-making[55]. This in turn has raised related concerns regarding a so-called function creep which occurs where data collected for one purpose is instead used for another[56]. An example of such would occur where data collected for performance review would instead be used to decide issues such as worker promotion or compensation awards, or potentially even for hiring and firing decisions. It is imperative to consider this data usage concern together with the heightened security concerns that come hand in hand with the increased exposure of employee data in the digital realm, whereby there is an exponential increase in workers’ personal data becoming endangered by issues such as data-leaks, cyber-hacking, or even to be misplaced through internal risks of bugs or other malfunctions to the automated decision-making AI software[57].

Chapter 3 Anti-Discrimination and Equality Laws

[55] Sandra Wachter ‘Affinity Profiling and Discrimination by Association in Online Behavioural Advertising’ (2020) 35(1) 367- 427 *Berkeley Technology Law Journal* https://pa- pers.ssrn.com/abstract=3388639

[56] European Parliament ‘Surveillance & monitoring: The future of work in the digital era’ (2020) https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656305/EPRS_STU(2020)656305_EN.pdf at p.3 and p. 22

[57] Müller, N., D. Kowatsch and K. Böttinger (2020), “Data Poisoning Attacks on Regression Learning and Corresponding Defenses”, *Proceedings of IEEE Pacific Rim International Symposium on Dependable Computing, PRDC*, Vol. 2020-December https://arxiv.org/abs/2009.07008v1 at pp. 80-89,

EU jurisprudence contains a body of primary[58] and secondary[59] laws that confer anti-discrimination and equality rights to workers who are from or associated to a class sharing protected characteristics, which operates on a tight taxonomical distinction between direct or indirect discrimination. The orthodox view is to conceptualize AI discrimination as indirect discrimination. The corollary of this however is that employers and platforms are open to defend and justify the use of potentially discriminatory algorithms. Alternatively, some algorithms may be directly discriminatory, but this will be practically rare and most often be confined to algorithms that are used to mask prejudice human decision-making.

Similarly, anti-discrimination laws in the US distinguish against two types of discriminatory treatment in the workplace: disparate treatment and disparate impact[60]. These are collectively established under Title VII of the 1964 Civil Rights Act[61], the Americans with Disabilities Act (ADA)[62] and the Age Discrimination in Employment Act (ADEA)[63]. From a conceptual point of view, disparate treatment can address discrimination arising from proxy and redundant encoding in algorithms. However, the need to prove a discriminatory intent, whether implicit or explicit, will exclude most algorithms from the classification of disparate treatment and require them to be dealt with under the law of disparate impact.

[58] Within primary law, one of the most important provisions for the protection of the workers from discrimination can be found in Article 21 of the Charter of Fundamental Rights (CFRU), which prohibits against a particular set of protected characteristics defined in the act as: "*sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation.*" Article 52(3) provides that the level of protection given under the CFRU shall be equivalent to those available under the European Convention of Human Rights (ECHR) In relation to the latter, Article 14 be will of relevance to issues of potential discrimination by AI in the workplace. The problem with CFRU Article 21 for the workplace context is that subject to the dictum in *Egenberger* (C-414/16) and *Mangold* (C-144/04) that alludes to direct horizontal effect, it is still not thought to have application to private parties.
[59] The secondary laws relevant to the workplace context are the Racial Equality Directive (2000/43/EC); the Gender Equality Directive (recast) (2006/54/EC); the Gender Access Directive (2004/113/EC); and the Employment Directive (2000/78/EC).
[60] Ricci v DeStefano [2009] 557 US 557
[61] 42 U.S.C. § 2000e-2
[62] 42 U.S.C. § 12101
[63] 29 U.S.C. §§ 621–34

Under both the law of indirect discrimination and disparate impact, employers may potentially defend the use of algorithms. This raises important issues of what types of reasons and uses of algorithmic and automated decision-making will be considered defensible. US law is more liberal on the available defenses to disparate impact claims than EU law, and therefore appears to accept business related reasons as sufficient to excuse the use of potentially discriminatory algorithms. In contrast, EU courts focus less on the business imperative and more on the legitimate objective behind the measure and therefore appear to not only require an excusable but a justifiable reason for the algorithm.

I. European Anti-Discrimination and Equality Laws

I.A. Direct Discrimination

Direct discrimination under European law defines a situation where one person is treated less favorably than another is, has been, or would be treated in a comparable situation on the basis of one of the protected characteristics defined in the relevant directives[64]. To illustrate such in a workplace is extremely problematic because of the intersectional and proxy-based nature of algorithmic discrimination as well as the fact that these processes are correlative rather than causative.

The historical refusal of the European Court of Justice (hereafter ECJ) jurisprudence to recognize intersectional discrimination[65] is problematic because of the multidimensional and dynamic process in which algorithms differentiate groups of persons from each other on the basis of many different, and often unknown, proxy variables, particular in ML systems. The

---

[64] Article 2(2)(a) Directive 2000/43/EC; Article 2(2)(a) Directive 2000/78/EC; Article 2(a) Directive 2004/113/EC and Article 2(1)(a) Directive 2006/54/EC

[65] Dagmar Schiek 'Intersectionality and the notion of disability in EU discrimination law' (2016) 53(1) 35-63 *Common Market Law Review* https://pure.qub.ac.uk/en/publications/intersectionality-and-the-notion-of-disability-in-eu-discriminati

two leading cases on this issue are *Parris*[66] and *Z*[67]. In both cases the ECJ refused intersectional discrimination and instead endorsed the need for applicants to prove, either one, or otherwise multiple basis of protected characteristics as the reason for the disparate treatment[68]. Such excludes AI discrimination from the ambit of direct discrimination. Even where AI discrimination against a protected characteristic has occurred, proof of such is often unattainable. Recognizing, intersectional discrimination would have cured the procedural difficulties which workers would face in such instances, as noted by Attorney General Kokott in *Parris*[69] in their opinion[70].

AI proxy discrimination is often analogical to associative discrimination. The latter concept was recognized in *Coleman*[71] where the ECJ held in context of a claim brought by a mother for discrimination against her employer by their refusal to provide time off in order for her to care for her disabled child that *"the prohibition of direct discrimination [...] is not limited only to people who are themselves disabled"* but also includes associated persons, e.g. parents and caregivers. In contrast to the ECJ's finding on intersectional discrimination, this jurisprudence considerably extends the scope of persons who come within the category of protected persons. This has potential utility for proxy and intersectional discrimination provided that there is sufficient proximity within the algorithmic correlation between the discriminatory act and the protected characteristic and that such can traced through the algorithm in order to succeed in a claim for discrimination[72]. This may be problematic in certain

[66] Case C-443/15 *David L. Parris v Trinity College Dublin* [2016]
[67] Case C-363/12Z *Z. v Government* [2014]
[68] Raphaele Xenidis, 'Multiple discrimination in EU anti-discrimination law: towards redressing complex inequality?' (2018) in Uladzislau Belavusau and Kristin Henrard (eds), *EU anti-discrimination law beyond gender* (Hart Publishing 2018), at p. 59 and p. 72.
[69] Case C-443/15 *David L. Parris v Trinity College* [2016] Opinion of Advocate General Kokott at para 4
[70] Case-152/11 *Johann Odar v Baxter Deutschland GmbH* [2012] at para 69; and Case C-312/17 *Surjit Singh Bedi v Bundesrepublik Deutschland* [2018] at para 75.
[71] Case C-303/06 *Coleman* [2008] at para 56
[72] Case C-668/15 *Jyske Finans A/S* [2017] at paras 16-21

types of predictive profiling or inferential analytics where there is some but not necessarily sufficient correlation between the algorithm and the protected characteristic[73].

Unlike US law, it is not necessary to prove a discriminatory intent in a direct discrimination claim[74]. This lack of an intention requirement in EU anti-discrimination law is advantageous when dealing with AI algorithms since these operate on pre-programmed patterns that are incidental rather than intentional[75]. EU law therefore examines whether the particular treatment occurred 'because' of' a protected characteristic, irrespective of the motivation of the defendant.

Prassl, Binns and Lyth[76] therefore argue that direct discrimination can be attributed to automated decisions where there is evidence that a discriminatory criterion has been applied by the algorithmic process that either targets a particular person, either directly or by proxy, because of their protected characteristics. Even with such argument, the chances of establishing a successful case of direct discrimination are limited since the doctrine still requires sufficient causality between the differential treatment and the protected characteristic, which reintroduces the problems EU law faces in regards to intersectional and proxy-based discrimination. To give an example in a case where direct proxy discrimination could succeed; if an employer creates an algorithm that seeks to let go of all pregnant woman as a means of avoiding costs of paying for maternity leave, such will amount to a case of direct proxy discrimination on the basis of sex[77], provided that proof of the underlying basis to directly discriminate against pregnant women can be adduced. Likewise, where search and click data has been collected by an online platform in regards to its user's online activity, and revealed a search history of a user looking for restaurants with wheelchair access, with the effect that the algorithm has then

[73] Xenidis, n 35 at pp.747-748
[74] *James* v *Eastleigh Borough Council* [1990] UKHL 6, [1990] 2 AC 751 per Lord Bridge of Harwich at pp.5-6
[75] Stephanie Bornstein, 'Antidiscriminatory Algorithms' *Alabama Law Review* (2018) 70(1) p.520,
[76] Jerimias Prassl, Reuben Binns and Aislinn Kelly-Lyth 'Directly Discriminatory Algorithms' (2022) *Modern Law Review*, (forthcoming)
[77] See e.g. C-177/88 *Elisabeth Johanna Pacifica Dekker*

as a result of this categorized the user as being disabled, when they themselves are not. If the search results are then tailored on the basis of this misclassification, such person would be denied opportunities and could accordingly be subject to direct discrimination[78].

However, the same scenario may not work where there is insufficient connection between proxy and protected characteristic. For instance, in the online platform example such could occur where the website had used other click data such as searches for car parking spaces, which would be harder to affiliate with disability unless the person had explicitly searched for 'disabled car parking spaces'. Coming within the category of direct discrimination will therefore involve significant causality between the basis of discrimination and the algorithmic process, which will most likely not be satisfied in cases where inferential analytics and regressions have been used to automate decisions[79]. Even if algorithms are not creatures of intention, they are also not creatures of intuition but of iteration. Accordingly, the orthodox view is that such kind of AI discrimination does not ordinarily but exceptionally constitute direct discrimination where there is a strong and disprovable link between a protected characteristic and a proxy[80]. It therefore remains that many, if not most cases, of AI discrimination in automated decision-making are more appropriately dealt with as indirect discrimination claims.

I. B. Indirect Discrimination

Indirect discrimination occurs where a facially neutral provision, criterion or practice puts protected groups at a disadvantage to others, unless that provision, criterion or practice is

[78] European Commission 'Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law' (2021) https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1 at p.68
[79] Xenidis, n 35 at pp.747-748
[80] Phillip Hacker, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law' (2018) 55(4) 1143-1152 *Common Market Law Review* https://ssrn.com/abstract=3164973

objectively justified[81]. The legal focus is therefore on the effects rather than the objective of the algorithm. Intersectional and proxy-based forms of differentiation are therefore more easily classed as instances of indirect discrimination. However, the corollary is that algorithmic discrimination is not per se unlawful and may be considered a justified practice of decision-making in certain situations.

The operation of indirect discrimination can be seen in the decision of the Bologna Labor Court in *Filcam VGIL Bologna and others v Deliveroo Italia SRL* [82] where the use of AI as a means of allocating work was found to be indirectly discriminatory. In this case, Deliveroo deployed an algorithm, known as Frank, to distribute the work slots of drivers on a priority system. The priority system was an automated process through which the AI calculated two 'scores' that were awarded to drivers based on (i) their reliability index, and (ii) their peak participation index. The system adversely impacted the drivers ability to obtain work since those who were awarded a lower priority subsequently had a lesser chance at obtaining a delivery slot, which was found to amount to indirect discrimination. The case affirms the orthodox approach of categorizing AI as indirect discrimination.[83] For one, intersectional and proxy-based discrimination in algorithmic and automated decisions are more easily treated as cases of discrimination. In doing so, the doctrine relieves the claimant from issues of proof since it is no longer necessary to dive into the black box phenomenon of the algorithm to detect how exactly the AI has acted discriminatory. Rather, it is sufficient to look at the effects of the algorithm and assess whether persons sharing a protected characteristic are disadvantaged when compared to others who do not share a protected characteristic[84].

---

[81] Article 2(2)(b) Directive 2000/43/EC; Article 2(2)(b) Directive 2000/78/EC; Article 2(b) Directive 2004/113/EC; Article 2(1)(b) Directive 2006/54/EC

[82] Court of Bologna, RG 2949/2019, ord. 12.31.2020,

[83] Xenidis, n 34 at p.747

[84] Toon Calders and Indre Zliobaite, 'Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures' in Toon Calders and others (eds), *Discrimination and Privacy in the Information Society* (Springer 2013) at pp. 52–53

Hacker argues that liability for a indirectly discriminatory algorithm may be elided because of the role of objective justifications in indirect discrimination claims[85]. The idea is that defendants could argue that the algorithms used serve legitimate purposes and are therefore lawful provided that they are neither disproportionate nor unnecessary to achieving their legitimate objective. This relates back to the technical trade-off in algorithms between accuracy and explainability and introduces a new dimension of efficiency and lawfulness[86]. These concerns materialized in the English Administrative Court decision in *R (The Motherhood Plan) v Her Majesty's Treasury*[87] where it was held that the Government was able to justify the potential discrimination by AI on the basis that the automation process was quicker, cheaper and more straightforward than the use of human decision-making. Arguably however, the threshold for justifying potentially discriminatory AI is high. Evidently, Deliveroo was not able to persuade the court in *Filcom v Deliveroo* that the use of AI was commercially justifiably. Likewise, in an Article 8 ECHR context, the decision in *SYRI*[88] by the Hague District Court also suggests that courts will not be too readily persuaded that the burden of justifying the use of a potentially discriminatory algorithm is easily discharged. It is therefore crucial to note that although the available defenses in EU law in relation to indirect discrimination appear broader than their US counterparts, they are actually more restrictive in effect. This can be seen by the fact that arguments relying on purely budgetary or financial justifications will be insufficient to trigger exemption from liability[89]. Rather, EU law seeks to ensure that any action that may differentiate between persons must comply with the substantive philosophy of anti-discrimination law in creating equal outcomes for all persons operating in the labor market. This contrasts with the philosophy of anti-discrimination law in the US which adopts a more

[85] Phillip Hacker, n 80 at p.1153
[86] European Commission, n 78 at pp.73-75
[87] *R (The Motherhood Plan) v Her Majesty's Treasury* [2021] EWHC 309 see paras 124-134
[88] Rechtbank Den Haag (C-09-550982-HA ZA 18-388) ECLI:NL:RBDHA:2020:865
[89] Joined Cases C-4/02 and C-5/02, *Hilde Schönheit v. Stadt Frankfurt am Main and Silvia Becker v. Land Hessen* at para 85

laissez faire approach that is focused on equality of opportunity rather than outcome[90], which is reflective in the greater willingness of courts to justify the use of differential practices on the basis of commercial and business needs than in Europe. The result is that EU law appears unlikely to accept the existence of mere statistical correlation as sufficient to justify the use of an algorithmic model where such would otherwise lead to a discriminatory outcome. Rather the courts will carefully examine the objective, necessity and proportionality of the algorithm to see whether it makes the decision just. Particularly the emphasis on necessity and proportionality will require courts to grapple with the question whether the use of the algorithm as a whole is necessary and proportionate to its objective, or whether it is just the use of certain datasets, target variables or processes that need to be proportionate and necessity or whether these can be rectified in such a way as to legitimize all other aspects of the algorithm.

II. US Anti-Discrimination and Equality Laws

II.A. Disparate Treatment

The conceptualization of disparate treatment doctrine is both broader and narrower than its European equivalent of direct-discrimination due to its greater acceptance to discrimination by proxy and association, as well as its reliance on proof of intention, whether explicit or implicit. The difference is further evident in the fact that US courts may award compensatory and punitive damages for disparate treatment[91].

In *Hodgson v. Approved Personnel Services* 1975 the Fourth Circuit ruled that the terms *"recent graduate"* in a job advertisement deterred older workers from applying in violation of

[90] Risa Lieberwitz 'Employment Discrimination Law in the United States: On the Road to Equality?', in R. Blanpain (ed.), *New Developments in Employment Discrimination Law* (2008) at p.5
[91] 42 U.S.C. § 1981a(a)(1)

the ADEA[92]. Evidently, an algorithm that recruits job applicants on the basis of experience may attract scrutiny for discriminating against younger applicants even though it operates on a facially neutral measure, which means that an employer in the United States may be prima facie liable for intersectional discrimination[93]. This facilitates the legal categorization of proxy and intersectional discrimination since courts are willing to recognize such representational dimensions of protected characteristics, provided that the plaintiff can then establish that the defendant intended to discriminate against such, which then challenges whether the law will also respond to non-causal, iterative and regression based operations of discriminatory treatment.

Despite the broader acceptance of allowing discrimination by proxy or association, it will still be necessary for the plaintiff to show a discriminatory intention[94], which may be extremely difficult where proxies are used, as can be seen by the court's decision in *Boyd v. City of Wilmington 1996*[95]. Accordingly, the ability to establish disparate treatment will depend on a positive finding of a discriminatory intention of a human operating behind the AI[96]. Kim[97] gives the example of where AI is used to mask an employer's underlying intention to discriminate, concluding that such will certainly constitute discriminatory intent. In such case, liability can be attributed from the hidden discriminatory intent of the employer behind the algorithm. Where there is no express evidence of discrimination, courts adopt a burden-shifting framework to evaluate a summary judgement motion, as established in *McDonnell-Douglas v. Green* where the court allowed for the burden to be shifted from plaintiff to defendant in the establishment of the discriminatory intention. Similarly, in *Price-Waterhouse*

---

[92] Hodgson v. Approved Personnel Serv., Inc., 529 F. 2d 760, 766 (4th Cir. 1975); Reid v. Google, Inc., 50 Cal.4th 512 (Cal. 2010)
[93] Lam v. University of Hawaii, 164 F.3d 1186 (9th Cir. 1998)
[94] EEOC v. Horizon/ CMS Healthcare Corp. 2020 F.3d 1184, 1191-92 (10th Cir.2000)
[95] Boyd v. City of Wilmington, 943 F. Supp. 585, 587, 590-91 (E.D.N.C. 1996)
[96] Charles A. Sullivan 'Employing AI' (2018) 63(3) 395-430 *Villanueva Law Review* https://digitalcommons.law.villanova.edu/vlr/vol63/iss3/2
[97] Kim, n 37 at pp.884-885

v. Hopkin the court allowed for a composite approach to the evidentiary analysis of intentional discrimination[98], which extends the scope of disparate treatment from cases only involving express intent to cases involving unconscionable bias, provided that such motivated the discrimination[99]. This approach is often known as the 'anti-stereotyping' theory and has been endorsed by the Supreme Court in *Los Angeles Department of Water & Power v. Manhart*[100]. That means that an employer may be found liable for discrimination on the basis of disparate treatment even where there is no deliberate intention to discriminate against a protected class, so long as the plaintiff is able to prove the underlying bias behind the algorithm. To give an example, if an employer used a screen algorithm to make hiring decisions which operated on a skew set of training data that was influenced by the employer's stereotyping when selecting the dataset, such could amount to disparate treatment in theory, even without deliberate intent. The advantage of classing AI discrimination under this heading is that the law of disparate treatment is unjustifiable and the employer will be unable to defend or excuse his discriminatory conduct. The disadvantage is that it will be difficult to find the hidden bias or intention behind the algorithm that is necessary for the attribution for liability under disparate treatment. Specifically in contexts of ML, it will be factually onerous to pinpoint the origin of stereotyping. In these situations, it may be easier from a procedural point of view to evidence discrimination on the basis of the output rather than the input of the algorithm.

[98] McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973); Price Waterhouse v. Hopkins, 490 U.S. 228 (1989). *See further* Desert Palace, Inc. v. Costa 539 U.S. 90 (2003); *See e. g.*, *Arlington Heights v. Metropolitan Housing Dev. Corp.*, 429 U.S. 252, 265-266, 97 S.Ct. 555, 563-565, 50 L.Ed.2d 450; *see also Kimble v. Wisconsin Dep't of Workforce Dev.*, 690 F. Supp. 2d 765, 778 (E.D. Wis. 2010)

[99] Note that there are occasionally higher standards such as in context of the ADEA where it is necessary not only to show that the protected characteristic was a motivating factor but a 'but-for cause' – see the decision in Cramblett v. McHugh, No. 3:10-CV-54-PK, 2012 WL 7681280, para 18; *see also* Jessica M. Scales 'Tipping the Balance Back: An Argument for the Mixed Motive Theory under the ADEA' (2010) 30(1) 229-262 *St. Louis University Public Law Review* https://scholarship.law.slu.edu/plr/vol30/iss1/11; Leigh A. Van Ostrand 'A Close Look at ADEA Mixed-motives Claims And Gross V. FBL Financial Services, Inc' (2009) *Fordham Law Review*; Ann Marie Tracey '*Still Crazy After All These Years? The ADEA, the Roberts Court, and Reclaiming Age Discrimination as Differential Treatment*' (2009) 46(1) 607-661 *American Business Law Journal* https://doi.org/10.1111/j.1744-1714.2009.01087.x

[100] City of Los Angeles v. Manhart, 435 U.S. 702 (1978)

II.B. Disparate Impact

Alternatively, AI discrimination may be categorized as disparate impact where a facially neutral employment practice produces a discriminatory effect on a protected group when applied[101]. The law of disparate impact is often described as an objective legal enquiry and therefore considered as a more apt legal remedy for issues involving systemic discrimination that are often more subtle to detect[102].

US courts often establish disparate impact by reference to an evidential threshold known as the '*four-fifths'* rule[103]. Although this threshold is designed as a guideline, it does have an exclusionary effect since it would suggest that the use of AI automated decisions, would need to result in a selection rate that is less than 80% for persons who are members or otherwise associated to a protected group[104]. In *Coleman v Exxon*[105] the court did not accept the evidence presented by the plaintiff that they had incurred discrimination by the use of an algorithm to rank individual employees and summarily dismissed the claim. Significantly to the operation of algorithms and data processing will also be the issue of sample size. Naturally, the threshold for meeting the four-fifths rule will be more severe for claimants who are part of a small sample size than those who are part of a greater sample since any statistical deviation between a protected class and others will be greater in the latter situation because of the sheer sample size[106].

---

[101] Barocas and Selbst n 33 at pp.701-712
[102] Griggs v Duke Power 401 U.S. 424, 432 (1971)
[103] Barocas and Selbst n 33 at p. 701
[104] Edward J Moreno 'Disability Bias Should Be Addressed in AI Rules Advocates Say' *Bloomberg* (New York May 6 2022) https://news.bloomberglaw.com/daily-labor-report/disability-bias-should-be-addressed-in-ai-rules-advocates-say at para 6
[105] Coleman v. Exxon Chem. Corp., 162 F. Supp. 2d 593, 609–11 (S.D. Tex. 2001)
[106] *See for general explanation* Daniel L. Rubinfeld, Reference Guide on Multiple Regression, *Reference Manual on Scientific Evidence,*at p. 189- 194, p. 214

Framing the legal enquiry in evidential terms has led commentators to question whether the law of disparate impact really provides an independent substantive framework for the identification of systemic discrimination or merely serves as an evidentiary tool on which intentional discrimination can more easily be established. This would mean that the law of disparate impact is a normative extension, rather than an alternative, to the law of disparate treatment[107].

This normative distinction is relevant to the ability of employers to audit AI decision-making, particularly in light of the Supreme Court's landmark decision in *Ricci v. DeStefano*[108]. The case concerned a claim brought by white and Hispanic firefighters against the City of New Haven for a decision made by the city not to certify the test that served for the promotion of firefighters to Lieutenant and Captain. The city's decision not to certify the test results was based on the fact that doing so would have led to a disproportionate number of white candidates being promoted in comparison to Hispanic candidates. The white candidates therefore argued that the city's disregard against their test results discriminated against them on the basis of race. The Supreme Court agreed with the claimants and found the action *"impermissible under Title VII unless the employer can demonstrate a strong basis in evidence that, had it not taken the action, it would have been liable under the disparate-impact statute*[109]." Kroll argues that if an employer audits a predictive algorithm that is otherwise discriminatory in its decision-making outcome, he may open himself up to a lawsuit for disparate treatment on the authority. In his mind, the court's decision in *Ricci* blurs the lines between disparate treatment and impact in such a case[110]. Kim disagrees with this reading of *Ricci* and instead maintains that the law not only necessitates but in fact encourages employers to perform an audit on an algorithm where

[107] Richard Primus 'Equal Protection and Disparate Impact: Round Three' (2003) 117(2) 494-587 *Harvard Law Review* https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1526&context=articles
[108] Ricci v. DeStefano 557 U.S. 557 (2009)
[109] Ibid para 563.
[110] Joshua A. Kroll et al 'Accountable Algorithms' (2017) 165(3) 633-705 *University of Pennsylvania Law Review* https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3 at pp. 692-695

faulty.[111] Kim's argument is more consistent since otherwise employer's would face a discrimination paradox: where an employer continues to use a faulty algorithm, they would be liable for disparate impact, but where they would in fact audit the algorithm, they may face liability for disparate treatment. Kim's argument is also more consistent with the business necessity defense and the ability of the claimant to prove that there is a less intrusive way of achieving the purpose other than by the discriminatory algorithm. Otherwise, the law would in fact encourage will-full blindness by the employer and deter any corrective action such as discrimination aware data-mining or methods such as data preprocessing, model post-processing or even model regularization. Krolls' argument is however useful in illustrating the role of human intent behind the machine for situations involving disparate treatment and the fine line between this law and that of disparate impact.

The Courts allow employers to defend disparate impact claims where they can demonstrate that a particular practice is either "*job related*" or consistent with *"business necessity*"[112]. These defenses are much more commercially-orientated than the available defenses in EU law as well as reflective of the employment-at-will context in which they operate under US law. To succeed with a business necessity defense, courts will require empirical proof that a potentially discriminatory criterion accurately relates or predicts to job-performance[113].

Often US courts will request a criterion, construct or content validation study, which is although not strictly legally required a commonly used method of assessing whether a given practice is a valid measure for job performance[114]. In relation to criterion and construct validation studies, many algorithms, especially those that are predictive or data trained models,

---

[111] Pauline T. Kim 'Auditing Algorithms for Discrimination' (2017) 166(1) 188-203 *University of Pennsylvania Law Review* https://scholarship.law.upenn.edu/penn_law_review_online/vol166/iss1/10 pp.191-197
[112] 42 U.S.C., §2000e-2(k)(1)(A)
[113] El v. Southeastern Pennsylvania Transportation Authority, 479 F.3d 232 (3d Cir. 2007)
[114] See e.g. EEOC Uniform Guidelines on Employee Selection Procedures (1978) 29 C.F.R. §§ 1607.5, 1607.15-16

are essentially a self-validating criterion validation study themselves since they operate on the basis of uncovering statistical correlations[115]. Commentators such as Kim therefore conclude that any review by law would amount to nothing less than a tautology. Likewise Barocas and Selbst conclude that most algorithms will pass the validation study provided that they are job related[116]. This can be seen in *Morales v McKesson Health Sols., LLC*[117] where the court found that the use of algorithms in reaching a termination decision constituted a legitimate workplace practice and therefore was neither discriminatory nor reason for wrongful termination. Bornstein however argues that in such cases a court could request a content validation study since such conducts its assessments on the worker's performance on the actual tasks of a job rather than an abstract assessment of their suitability and characteristics as individuals and therefore catch discrimination that would otherwise be missed by the other validation studies, particularly under a criterion-based study[118].

Even if a defense is established, the plaintiff may still prevail by proving that the employer could have used an alternative employment practice with less discriminatory results that the employer refused to adopt[119], as noted that the US Supreme Court in *Albermale Paper C. v Moody*[120], which is similar to the doctrine of proportionality under EU law despite the reversal of the burden of proof. Barocas and Selbst argue that this alternative employment practice may require the rectification of the algorithm itself, which can be achieved by reverse engineering the process or removing the faulty dataset[121]. But crucially, the procedure behind this doctrine is very different when comparing EU and US law in regards to this state of the court's legal analysis. Under European law, the burden of proof will shift to the employer to

[115] Kim, n 37 and n 35 at p.866 and p. 908
[116] Barocas and Selbst, n 33 at p.709
[117] Morales v. McKesson Health Sols., LLC, 136 F. App'x 115,116 (10th Circ. 2005)
[118] Bornstein, n 75 at pp.565-567
[119] § 2000e-2(k)(1)(A)(ii).
[120] Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975)
[121] Barocas and Selbst, n 33 at p.705 onwards

provide an acceptable defense. In contrast, the burden of proof in the US remains with the plaintiff once an employer has articulated a non-discriminatory reason for the use of the contested measure. Disproving the employer's claim in relation to the AI may be very complicated to say the least absent of the worker having substantial technical knowledge or otherwise assets to allow them to obtain such through expert evidence.

Chapter 4: Privacy and Data Protection Laws

Privacy has historically played a significant role in the core legal infrastructure of the EU and has always been strictly regulated, as evident in the landmark enactment of the GDPR (General Data Protection Regulation)[122]. In contrast, privacy rights are historically less regulated and more fragmented in the US[123] until the recent piecemeal legislative shift to adopt privacy and data protection measures, which can be seen in the emergence of incoming laws in California, Colorado[124], Connecticut[125], and Virginia[126], as well as in the proposal of the American Data Privacy Act (ADPA)[127] by the federal government. Unlike the GDPR, most of these privacy laws, except for the incoming law in California, namely the California Privacy Rights Act (CPRA)[128], do not include employees, workers or job applicants within the scope of their regulatory regime. The research proposal will therefore predominantly analyze the

---

122 Regulation (EU) 2016/679 (GDPR)
123 Ifeoma Ajunwa, Kate Crawford & Jason Schultz, 'Limitless Worker Surveillance' (2017) 105(1) 736-776 *California Law Review* https://29qish1lqx5q2k5d7b491joo-wpengine.netdna-ssl.com/wp-content/uploads/2017/07/3Ajunwa-Schultz-Crawford-36.pdf
124 Colorado Privacy Act (CPA) S.B. 21-190 (Col. 2021)
125 Connecticut Data Privacy Act (CTDPA), S.B. 6, 2022 Gen. Assemb., Reg. Sess. (Conn. 2022)
126 Virginia Consumer Data Protection Act (VCDPA) S.B. 1392 (Va. 2021).
127 American Data Privacy and Protection Act (ADPPA) H.R. 8152, 117th Cong. (2022)
128 The California Privacy Rights Act (CPRA) was passed as a ballot initiative in California in 2020 as Proposition 24. https://www.oag.ca.gov/system/files/initiatives/pdfs/19-0021A1%20%28Consumer%20Privacy%20-%20Version%203%29_1.pdf.

California Draft Regulations[129] mandated under the CPRA, which provide significant rights to consumers, and employees, in relation to automated decision-making.

In the EU, the main right for such stems from Article 22 of the GDPR, which creates a *right not to be subject to solely automated processing decisions*[130]. There is also rich academic debate whether the GDPR provides a *right to an explainable automated decision,* but such right is less well established in practice. The issue with the GDPR is that the act is very strict as to what types of automated decisions come within its ambit, since it only addresses solely automated decisions, and also imposes a rather strict legal threshold that requires proof of a particular harm.

Likewise, it is also possible to carve out a *right not to be subject to automated decision* and a *right to explainable automated decisions* from the draft regulations under the CPRA. This would provide a superior level of regulatory oversight since the act applies to a broader variety of automated decision-making processes, without requiring proof of harm. These draft regulations also imposes stricter consent requirements than the GDPR. However, it is important to consider the draft regulations in context of the proposed ADPA. This is because the latter act doesn't include employees within its data protection remit and therefore may have the effect of dismantling some of the state level protections, if enacted, since the ADPA is subject to the federal presumption that it supersedes conflicting state laws.

I. EU Privacy and Data Protection Laws

[129] The Draft Regulations are available here: https://cppa.ca.gov/meetings/materials/20220608_item3.pdf.
[130] Diana Sancho 'Automated Decision-Making under Article 22 GDPR: Towards a More Substantial Regime for Solely Automated Decision-Making', Martin Ebers and Susana Navas (eds.) *Algorithms and Law* (Cambridge University Press 2020) pp. 136-156

I.A. General Data Protection Regulation

i. Rights Relating to the Automated Decision

To come within the protective ambit of Article 22, a *"decision"* has to be made by the applicable data processing operation, as opposed to merely a preparatory, supporting or complementary step made in anticipation of a decision[131]. Accordingly the provision does not permit workers to avoid any and all exposure to algorithmic processing.

Likewise, decisions will not come within Article 22 where there is some degree of human agency involved. The question then is; exactly what degree of human involvement is sufficient to disrupt the automated process? Wachter argues for a narrow interpretation where any human involvement, however *de mimis*, will disqualify Article 22[132]. In contrast, the Article 29 Working Group, argue that human agency must be meaningful rather than *de mimis*[133]. National authorities cautiously agree with the latter interpretation. However, the threshold for meaningful human agency is low, which is concerning in a labor law context where practically many, if not most, decisions made by AI will always have some degree of human involvement, even if the degree of activity is limited[134]. The Amsterdam court in *Über* held that the decision to deactivate a platform driver's license was not a "s*olely*" automated process since an Operational Risks Team made the final decision[135]. In contrast, where human

[131] Christopher Kuner et al. *The EU General Data Protection Regulation (GDPR) – A Commentary* (Oxford: OUP, 2020), pp. 530-532; *see more generally* Elena Gil González, Paul D. Hert, 'Understanding the Legal Provisions that Allow Processing and Profiling of Personal Data – An Analysis of GDPR Provisions and Principles' (2019) 19(1) 597-621 *ERA Forum*,

[132] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) *International Data Privacy Law*, 76–99 https://doi.org/10.1093/idpl/ipx005; *see also* Adrián Todolí-Signes 'Algorithms, artificial intelligence and automated decisions concerning workers and the risks of discrimination: the necessary collective governance of data protection' (2019) 25(4) 465-481 *Transfer: European Review of Labour and Research*,

[133] Michael Veale and Lilian Edwards 'Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling' (2018) 34(2) 398-404 *Computer Law & Security Review* https://doi.org/10.1016/j.clsr.2017.12.002

[134] Javier Sánchez-Monedero, Lina Dencik, and Lilian Edwards 'What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems' (2020) *ACM Conference on Fairness, Accountability, and Transparency*, 27–30, New York, ACM 11https://doi.org/10.1145/3351095.3372849 pp.458-468

[135] Note the alternative approach to find liability under Article 15 GDPR in Rechtbank Amsterdam (C/13/692003/HA RK 20-302) ECLI:NL:RBAMS:2021:1018; see further for similar approach by the Austrian

agency does not take place in the final decision but in the prior stages of leading up to such, such as by selecting the factors to be used by the algorithm, such use of AI will constitute a solely automated decision. By example, the Italian Data Protection Authority in *Foodinho* SRL[136] held that Article 22 was infringed by the use of an automated decision-making system called Jarvis, which used data profiling and processing activities to create an excellence-system to allocate work to platform workers. Even though the parameters of Jarvis were set by the employers at Foodinho, the algorithm amounted to a 'solely' automated decision because the final decision was taken by the algorithm not the employees. Evidently, where humans have no influence on the outcome, the consensus is that the algorithmic decision will come within the ambit of Article 22[137].

Where there is human agency with the final decision, courts will scrutinize whether such amounts to meaningful influence or is merely a rubber stamping process. This can be seen in the Data Protection Authority's decision in Portugal on the use of AI to predict student's behavior in exams through the processing of their biometric data through motion and facial detection when using school webcams[138]. The Data Protection found that the degree of human involvement of teaching staff was merely a confirmation process of a decision reached on automated grounds. Occasionally it will be difficult to draw clear boundaries between decisions that are and are not solely automated. For example, the Austrian Federal Administrative Court found that the AI processing of the allocation of funding to jobseekers did not amount to a

---

Data Protection Authorities in AEPD, Procedimiento No: PS/00477/2019, https://www.aepd.es/es/documento/ps-00477-2019.pdf; see also AEPD, Procedimiento No: PS/00500/2020, https://www.aepd.es/es/documento/ps-00500-2020.pdf

[136] Garante per la protezione dei dati personali (Italy) – 9675440 https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9675440

[137] The above approach under the GDPR consistent with pre-GDPR case law, as can be seen in when looking at the decision made by the French DPA on the use of algorithmic decision-making as a determinant to student admissions to universities where human staff had no influence in the final decision made on whether or not offers should be sent to applicants; see for Commission Nationale de l'informatique et des Libertés, Décision 2017-053 du 30 août 2017 www.legifrance.gouv.fr/cnil/id/CNILTEXT000035647959/

[138] CNPD (Portugal) - Deliberação/2021/622 https://www.cnpd.pt/umbraco/surface/cnpdDecision/download/121887

solely automated decision because a counsellor reviewed, and if necessary, diverged from the results of the algorithm[139]. In contrast, the first instance found that the counsellor's role was merely to confirm the decision and that any intervention made was therefore based solely on the findings of the automated decision[140]. These diverging perspectives on automated decisions may alternatively be reviewed through a Data Protection Impact Assessment[141]. This applies to automated decisions as opposed to solely automated decisions and comes into play where the AI constitutes a "*high risk*" to the personal data and the "*rights and freedoms*" of individuals.

It is necessary that the automated decision has "*legal or similarly significant legal effects*" to come within Article 22. This will be straightforward where a decision is binding or when it impacts the rights and obligations of that person, such as in hiring or firing, or promotion decisions. It is slightly less clear what decision amounts to "*similarly significant effects*", but it is thought that such can nonetheless be made out where the decision has the effects *"that are important enough to deserve attention and that significantly affect the conduct or choices of the person concerned* [..]"[142]. Useful guidance can be seen in the analysis of the Amsterdam Court in *Ola* and *Über* which considered, *inter alia,* the type of data subject to the automated decision, the immediate and long term consequences of the decision, the subjective impact of the decision on the data subjects. This imposes a harm threshold on the regulation of AI in the workplace so that the operation of AI and ML in automating decision-making is not *per se* unlawful but only so where it adversely impacts workers.

Additionally, and much the same as with indirect discrimination, there are defenses under Article 22(2), such as where the use of automated decision-making is necessary for the

[139] Bundesverwaltungsgericht, (Case W256 2235360-1/5E) ECLI:AT:BVWG:2020:W256.2235360.1.00,
[140] DSB-D213.1020, 2020-0.513.605
[141] Article 35(1)-(3) GDPR
[142] Rechtbank Amsterdam (Case C/13/689705 / HA RK 20-258) ECLI:NL:RBAMS:2021:1019 at para 4.51

entrance or performance of the contract[143]; or where such is authorized by EU domestic laws to which the data controller is subject, and which lays down suitable measures to safeguard the data's subjects rights, freedoms or legitimate interests[144], or is based on the data subject's explicit consent[145]. To give an example, how would courts approach the issue of whether the use of AI in a recruitment decision can be justified on contractual necessity – is it arguable that a hiring decision urgently requires AI to sieve out, what are often 100s of candidates, for a particular position so that the company can make an offer[146]. Indeed, particularly smaller companies with less employees and smaller budgets for human resources may argue that they otherwise would be inundated in the recruitment process[147]. It is then also arguable that automating decisions to AI, particularly in gig and platform contexts, is necessary for the performance of the contract and for the organization of labor power. Accordingly, this exception may not necessarily provide a get-away card in practice, but nonetheless a risk of a legal loophole for companies to defend their use of AI in automated decision-making. There is also the issue of consent and the fact that such is particularly precarious in an employment context where workers are subordinated to the employer and therefore not able to give free and informed consent.

Admittedly, the exceptions are subject to Article 9(1) safeguards[148] where the automated decision is based on sensitive categories of personal data, unless the decisions are based on reasons of substantial public interest, based on European or domestic laws, and suitable measures to safeguard the data subject's rights[149] and freedoms and legitimate interests

[143] Article 22(2)(a) GDPR
[144] Article 22(2)(b) GDPR
[145] Article 22(2)(c) GDPR
[146] see Cathy O'Neil *Weapons of Math Destruction: How Big Data Increases Inequality and threatens Democracy* (Crown Books: 2016) at Chapter 6
[147] Frank Hendrickx 'Privacy 4.0 at Work: Regulating Employment, Technology and Automation' (2019) 41(1) 147-172 *Comparative Labor Law & Policy Journal* https://cllpj.law.illinois.edu/access?returnurl=https://cllpj.law.illinois.edu/archive/vol_41/
[148] Article 22(4) GDPR
[149] Article 9(2)(a) GDPR

are in place[150]. In the aforementioned example, these safeguards will not feature much relevance. Even where the safeguards do not overrule the three exceptions, Article 22(3) provides that these shall be subject to suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

ii. Further Rights to the Explanation of the Automated Decision

A right to explanation is not only relevant to combatting transparency and accountability concerns in relation to the general use of personal data in profiling processes but particularly in relation to automated decision-making where workers are directly impacted by the outcome of the algorithmic assessment. There is therefore further debate whether Article 22 merely provides a negative right not to have a decision made without a human-in-the-loop or whether it extends to a positive right to an explanation[151]. Such reading would be carved out of a composite reading of Article 22(3), supported through Recital 71[152], the GDPR notification duties[153] and the access rights[154]. A right to explanation could have legal implications for the ability of workers to rely on labor protection laws such as the law of unfair or wrongful dismissal. Even so, the preferred view is that data subjects do not have an absolute right to explanation but rather a more specific right to be properly informed about particular aspects of the automated decision process[155]. These rights then enable workers to make use of other ancillary rights provided to them under the GDPR, such as the right to rectification[156] or to

[150] Article 9(2)(g) GDPR
[151] Lilian Edwards and Michael Veale.' Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking' (2017) 16(1) 18-84 *Duke Law & Technology Review* 16(1) 18–84 https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1315&context=dltr; Bryce Goodman and Seth Flaxman 'European Union regulations on algorithmic decision-making and a "right to explanation"' (2017) 38(3) 50-55 *AI Magazine* https://doi.org/10.1609/aimag.v38i3.2741 at p. 55
[152] Goodman and Flaxman, n 151 above at p. 53
[153] Articles 13-14 GDPR; Recitals 60-62
[154] Article 15 GDPR; Recital 63
[155] Wachter, Mittelstadt and Floridi, n 132 at pp.89-90
[156] Article 16 GDPR

erasure[157]. Additionally, workers may ask for their data processing to be restricted[158] or object to the data processing if such is in their employer's legitimate interests[159].

II. US Privacy and Data Protection Laws

II.A. California Privacy Rights Act

i. Rights Relating to Automated Decision Making

Particularly in light of the above legislative context, the agenda of the CPRA should be regarded as a landmark development in the protection of data and privacy rights and interests in context of automated decision-making. Unlike other state privacy laws, the CPRA explicitly refers to the regulation of privacy rights of employees in the workplace context and also applies to job applicants. This is especially noticeable since the predecessor of the CPRA, namely the California Consumer Privacy Act (CCPA)[160], operated on the opposite approach and contained an express 'employee-employer' exemption. However, this will be changed under the new regime of the CPRA from the 1st of January 2023.

The act creates draft regulations that provide specific legal rights pertaining to the use of AI in this context by granting consumers[161] with access and opt-out rights[162] to automated decision-making. Much like the GDPR, these draft regulations also respond to the use of automated and algorithmic decision-making by granting the right to object to such. It is interesting that both jurisdictions respond to the use of automated decision-making with opt-out rights rather than opt-in rights. The opt-out approach means that the use of automated

[157] Article 17 GDPR
[158] Article 18 GDPR
[159] Article 21 GDPR
[160] California Consumer Privacy Act (CCPA) 2018 (AB 375)
[161] The term "consumer" includes employees, applicants, independent contractors, and other types of workers.
[162] The CPRA adds these rights into § 1798.185(a)(16) California Civil Code

decision-making is prima facie lawful except where non-compliant with certain legislative conditions, which in this case will arise from a revocation of consent. In contrast, an opt-in approach would have meant that the use of automated decision-making is unlawful, except where the prerequisite consent has been obtained. This would have aligned such to many other legal contexts where consent acts as a gatekeeper to an otherwise unlawful procedure taking place. Contrary to such, an opt-out approach assumes that consumers, or in this case workers, have the power and capacity to initiate the opt-out and as such places the burden of obtaining redress on the worker to achieve meaningful change and contributes to the black-box problem in AI decision-making. This assumption is unrealistic to many employment conditions and difficult to reconcile with the fact that work contracts are premised on the existence of subordination rather than an equality of bargaining power between worker and employer.

It is interesting that the rest of the provisions within the CPRA and GDPR are a hybrid model between opt-out and opt-in procedures, since various other data processing requirements require affirmative consent. The decision of both legislators to identify automated decision-making as opt-out procedures is therefore revealing since it illuminates the legal presumption of them being lawful until consent is withdrawn. This suggests that they are considered to be on a lower hierarchy of digital harms in comparison to other issues that are presumptively unlawful unless they have consented. Significantly, and different to the GDPR, the draft regulations under the CPRA do not entail any exemptions to the opt-out right. This suggests that the CPRA places more emphasis on the consent criterion as an unavoidable ingredient to the use of automated decision-making.

The other key differences between the two regulations is that the current drafting of the CPRA has a much broader and more encompassing application than the GDPR and will therefore provide workers with both a broader and stricter set of protections under the draft regulations.

The jurisdiction is broader since the CPRA not only regulates solely automated decisions but all automated decisions. The CPRA therefore may provide redress to situations that may fall outside of the GDPR, such as where there is limited, but nonetheless some, degree of human involvement and agency. Indeed, this avoids situations of rubber-stamping or automation bias that can occur where humans often over-rely on the perceived mechanistic objectivity of AI to avoid responsibility for decision-making, and neglect their own personal experiences and value judgements[163]. It also demonstrates a more tightly regulated solution towards the need to have a human-in-the-loop in AI decision-making. Indeed, whilst workers only have the right not to be subject to a decision that is made completely absent of human agency under the GDPR, the CPRA arguably goes further and ensures that there is a right to opt-out of any automated decision-making process, even if there already is some human intervention. Potentially, this implies a right to not only have a human-in-the-loop but a right to have a decision made entirely by a human. If so, this would substantially exceed the legislative solution of the GDPR in the workplace since workers not only would have the right to have human management involved in the decision-making process but have a right to have the decision made entirely by humans.

The jurisdiction of the CPRA is also stricter because it applies independent of whether there are legal or similar effects. For instance, targeted job advertising would easily come within the scope of the CPRA. In contrast, it would not be captured by the GDPR because of the fact that it would be difficult to establish that such has a significant impact or legal effect on the job applicant without conducting a detailed legal investigation into the intrusiveness of the tracking process on the job applicant and the extent to which such had modified the individual's search, as well as a counterfactual analysis of how likely they would otherwise have been in obtaining neutral search results. With the CPRA, such analysis is not necessary

[163] Karppi, n 52 at p.2

for the CPPA to intervene since the act of conducting automated decision-making is per se subject to legal oversight, without need to prove adverse impact. In a sense, this suggests that the CPRA creates a system that is closer to absolute liability when compared to the GDPR.

ii. Further Rights to the Explanation of the Automated Decision

The CPRA also creates the right to know[164], the right to correct[165], the right to deletion[166], the right to limit use and disclosure of sensitive personal information[167], the right to opt-out of sales and sharing information[168], and the right to non-discrimination[169] in its draft regulations. Whether such amounts to a right to explanation in the sense that is envisaged by the academic literature in the EU will depend greatly on the strictness of enforcement of these rights, particularly the right to know. Under a strict textual reading, the right to know encompasses two distinct sub-rights. Namely, the right to an explanation of how the employer collects and handles personal information as well as the right to copies of "*specific pieces of personal information*". This therefore may not capture the right to have an explanation of the automated decision itself. In this sense, these rights are more concerned with providing a multifaceted protection of privacy in the workplace rather than a specific solution to automated decision making.

Even so, this multidimensionality will capture aspects of automated and algorithmic decision-making since ML, or other AI systems as well, often pool from vast amounts of datasets in order to operate. Specifically in the workplace, these data sets could be sourced from employee surveillance and tracking devices that monitor communications, language and

[164] The CPRA consolidates the right to know in §§ 1798.110–115 by expanding consumers rights to know certain personal information collected on or after January 2022 beyond the current 12 month period stipulated in §1798.130, subject to the proviso that such does neither imposes an impossible or disproportionate effect.
[165] This is a new right that the CPRA inserts into § 1798.106
[166] The CPRA consolidates the right spelt out in § 1798.105
[167] This is a new rights inserted by the CPRA into § 1798.121,
[168] The CPRA consolidates the provision of such under § 1798.120
[169] The CPRA consolidates the right spelt out in § 1798.125

behavior patterns between workers To give examples; many organizations currently deploy intrusive tracking devices to monitor employee's physical whereabouts, activity on their computers, keystrokes, and sometimes even biometric information such as heart rates or blood pressure[170]. From this perspective, it is not only the use of automated and algorithmic decision-making that raises transparency and accountability issues but also the collection of data itself and the act of monitoring workers[171]. Such uses of AI in the employment context not only create isolated instances of privacy violations but fundamental challenges to good data governance that will effect a multilayered system of transparency within AI decision-making

II.B. American Data Privacy Act

In light of the hierarchical relationship dynamics between state and federal laws, it is necessary to consider the fact that the ADPA excludes employees in its scope of covered data subjects as well as de-identified data, and data in the public domain[172]. This could in fact limit the protective scope of state national laws that do provide workers with privacy protections, since they would be deemed inconsistent with the federal instrument. If the ADPA would have included workers privacy protections, it would have created special requirements on certain types of covered entities and service providers that come within the act's definition of large data holders[173]. These special requirements include heightened obligations on transparency[174],

[170] Christopher Rowland 'With Fitness Trackers in the Workplace, Bosses Can Monitor Your Every Step—and Possibly More' *Washington Post* (Washington 19 February 2019), https://www.washingtonpost.com/business/economy/with-fitness-trackers-in-the-workplace -bosses-can-monitor-your-every-step--and-possibly-more/2019/02/15/75ee0848-2a45 -11e9-b011-d8500644dc98_story.html [perma.cc/JAJ6-S2DP]
[171] Ajunwa, Crawford & Schultz, n 123 at p. 741
[172] § 2(8) ADPA
[173] See § 2 (17) ADPPA; which defines large data holders as covered entities and service providers as entities with a gross annual revenue of at least USD 250 million who collect, process or transfer covered data of more than 5 million individuals, or alternatively of more than 200,000 individuals if such pertains to covered sensitive data (§ 2(24) ADPA)
[174] § 202 ADPA

individual's rights[175] and also an obligation to conduct a privacy impact assessment[176]. Significantly, large data holders are specifically mandated by the proposed act to conduct "*algorithmic impact assessments*"[177].

Chapter 5. Law Reforms in Europe and the United States

Both the US and the EU are in the process of enacting a series of specific and targeted legislative resolutions to AI technologies. These legislative reforms focus mainly on the issues of transparency and accountability within AI, but would also have utility for bias and discrimination issues by virtue of the interlocked nature of these general challenges posed by AI technologies. The reforms respond to the existing difficulties of taxonomizing the legal challenges raised by AI technologies within the existing framework of anti-discrimination and equality laws or privacy and data protection regulations respectively. These reforms recognize the idiosyncrasies of the legal issues raised by emerging AI technologies and create a blueprint for the regulation of accountable and responsible decision-making through algorithmic and automated processes.

I. Artificial Intelligence Act

On the 21st April 2021 the EU Commission introduced the AI Act (AIA)[178]. The proposed act and its Explanatory Memorandum[179] provide a landmark regulatory regime in

175 § 203 ADPA
176 § 301(d) ADPA
177 § 207(c) ADPA
178 European Commission 'Proposal for a Regulation of Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' (2021) COM (2021) 206 final https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206
179 European Commission 'Explanatory Memorandum for a Regulation of Rules on Artificial Intelligence (Artificial Intelligence Act)' (2021) https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN (hereafter Explanatory Memorandum)

response to the novel legal challenges raised by AI[180] that create liabilities for automated and algorithmic decision-making in the workplace when read together with the GDPR and EU anti-discrimination laws[181]. In contrast to the GDPR, the AIA is not a floor of rights but a ceiling, which prevents member states from legislating a higher standards of worker protections than those that are regulated within the proposed act[182]. Such makes it even more important that there is both adequate and suitable protections within the proposed act that can regulate AI and ML in the workplace.

AI technologies are defined in Article 3(1) and Annex I and ranked according to a risk-based approach. At the high end of the spectrum, the use of AI with an "*unacceptable risk*" is presumptively unlawful[183], and *"high risk"* AI may be unlawful where such is uncompliant with express regulatory conditions spelt out in the proposed act[184]. In contrast, the use of *"limited risk"* AI only is subject to transparency obligations[185], and "*minimal risk"* categories of AI are subject to voluntary rules[186]. Since the threshold for the unacceptable risk is extremely high[187], the use of AI in the workplace will more likely be classed as high risk.

Within this category, Annex II-A lists two previously recognized categories of high risk AI and Annex III establishes eight novel categories. The latter makes particular provision for the regulation of AI as high risk in matters concerning employment, workers management and access to self-employment. The act specifically details this category as applying to two

[180] It should be noted that the Act is by no means finalised since it is still subject to extensive negotiations between member states and therefore may undergo significant changes prior to its finalisation.
[181] EDPS 'Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)' (2021) https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf at p. 7
[182] Explanatory Memorandum para 2.2
[183] Title II and Article 5
[184] Title III
[185] Title IV and Article 52
[186] Title V and Article 69, Explanatory Memorandum at para 5.2.7
[187] The proposed act limits this category to 4 distinct uses of AI under Article 5(1)(a)-(d); *see further* Michael Veale & Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22(4) 97-112 *Computer Law Review International*
https://doi.org/10.48550/arXiv.2107.03721

situations which are; (i) where AI is used for recruitment or selection purposes, screening or filtering applications, evaluating candidates, (ii) as well as for making decisions on promotions and termination of work related contractual relationships, for tasks allocation and for monitoring and evaluating performance and behavior of persons in such relationships[188]. Algorithmic and automated decision-making in the workplace therefore come within scope of the act and therefore are prima facie covered.

However, the Explanatory Memorandum states that *"the classification as high-risk does not only depend on the function performed by the AI system, but also on the specific purpose and the modalities for which that system is used"*[189]. This is further supported by the emphasis the act plays on the intentions of the user and provider when analyzing the categorization of risk. Indeed, this distinction between providers, which is particularly significant to the employment context where users and providers may not be one and the same person. For one, the. act therefore places most of the regulatory obligations on the former, even though the latter arguably has more control over the deployment of the AI[190]. More importantly, the high risk obligations in the AIA are only applicable where the AI is used in accordance to the provider's intended purposes[191] for recruitment or selection of natural persons or for making decisions on promotion, termination, task allocation or monitoring and evaluation of performance and behavior of workers. The consequence of this is that liability may be evaded where a user has different intentions with the AI to those of the provider. Such can be seen in a recent illustration of Über, where the AI was intended by the provider to be used to predict

[188] Annex III, 4
[189] Explanatory Memorandum at para 5.2.3
[190] Aislinn Kelly-Lyth, 'The AI Act and Algorithmic Management' (2021) Dispatch No. 39 (European Union), *Comparative Labor Law & Policy Journal* (forthcoming)
[191] Article 3(12) AIA

safety incidents, and instead was used by the user as a means of determining whether sanctions should be issued to drivers[192].

The distinction placed on the differentiation of the providers' and users' intention raises significant accountability issues in a workplace context. Similar to a 'computer said so defense[193], the AAA, effectively implies a 'provider instructed so' defense, since the liability for the AI depends on the classification of risk, which in turn is mostly dependent on the intentions of the provider. Accordingly, a lack of liability may arise from the disparity between actual use and intended use of the AI in a algorithmic or automated decision-making context.

Accordingly the main legislative thrust of the act falls onto the provider rather than the ultimate user, who will be the employer in the workplace context. What then must the provider do and how can this be useful to the protection of worker's rights? The central obligations for high-risk AI are found in Chapter 2 of the AIA. The first are design criteria that are implemented at a pre-market stage under Article 9. This includes the establishment of a risk-management system that amounts to a continuous iterative process. This also includes the obligations in Article 10 that ensure that data sets are subject to good data governance and management practices. There are also obligations of data documentation[194], record keeping[195], transparency and provision of information to users[196], and of human oversight[197], as well as obligations relating to accuracy, robustness and cybersecurity[198].

Specifically the legal requirement that there should be human oversight in Article 14 AIA is a very interesting legislative development in dealing with human-in-the-loop issues in automated and algorithmic decision-making. Again, the focal point of the provision is similar

---

[192] Miriam Kullmann, Miriam and Aude Cefaliello, 'The Draft Artificial Intelligence Act (AI Act): Offering False Security to Undermine Fundamental Workers' Rights' (2021) http://dx.doi.org/10.2139/ssrn.3993100 at p. 5
[193] Karppi, n 52 at p.2
[194] Article 11 AIA
[195] Article 12 AIA
[196] Article 13 AIA
[197] Article 14 AIA
[198] Article 15 AIA

to the act as a whole since it is more concerned with the "*design and development"* of human oversight tools rather than the actual exercise of human oversight. It therefore provides for the existence of "*appropriate human machine interface tools*" but does not guarantee that these tools are in fact used[199]. The onus is therefore again on the provider. It is nonetheless likely that the practical effect of the act will necessitate compliance with the user in order for the provider to fulfil this legal obligation, especially if this provision is read together with Article 29(1) which requires users to comply with the instructions of the provider. Even so, it begs the question whether the act satisfies the need for a human-in-the-loop in a workplace context where it may not be sufficient to have anyone in the loop but rather it will be necessary for workers to feel that there is in fact not only a technically qualified person in the loop pursuant to Recital 48 of the AIA but someone who is qualified as a manager who understands the labor-related subject matter on which the AI is basing its decision.

Admittedly, the AIA does require providers to factor in that the user may "*foreseeably misuse*"[200] the AI, and to inform users where the AI may "*lead to risks to the health and safety or fundamental rights"*[201]. This builds on the obligation for the provider to take into account the environment for which the AI is to be used[202] but it does not necessarily guarantee that all obligations flow from the provider to the user. Notably, the provider is required to communicate the potential risks of the AI to the user of the user[203]. The corollary of this would be that the employer should consider the provider's risk management assessment when implementing the AI tool at work and carry out due and proportionate post-market surveillance assessments on the AI. The employer does have their own set of obligations concerning the accuracy of data and the avoidance of potential discrimination[204]. It should be noted that the

[199] Article 14(1) AIA
[200] Article 9(2)(b) AIA
[201] Art 13(3)(b)(iii) AIA
[202] Article 9(4) AIA
[203] Article 13(3)(iii) AIA
[204] Article 15(3)-(4); see also Article 29(1)-(5) AIA

Explanatory Memorandum makes express provision for the fact that the AIA should be read in conjunction with the existing social acquis of EU law, which will require the application of the act to have due regard to the aforementioned anti-discrimination and data protection laws[205].

The other noteworthy feature of the act that may face resistance in a workplace context can be seen in the fact that the act is comprised of an extremely principled and broad regulatory agenda that may be difficult to apply to pragmatic and specific situations in the workplace context. This relates to the fact that the AIA adopts a very function-centric rather than harms-centric focus of AI technologies. This is evident in the categorization of risks within the annexes that draw significantly upon European product liability laws. The issue with such is that it presupposes a correlation between liability and uses rather than liability and harms. But such correlation may not necessarily be present in the labor context. It also elides the fact that very small reconfigurations of algorithmic processes can in fact lead to huge differences in their outcomes. For instance, if an employer changes one of the variables of a data set, such as by requiring the algorithm to place greater emphasis on experience when recruiting candidates, the employer may be excluding an entire category of applicants since experience can be a proxy for age discrimination.

II. Emerging Federal and State AI Regulations

The US is currently in the process of considering a federal solution in form of a proposed Algorithmic Accountability Act (AAA), which was introduced as a bill to both Houses of Congress in February 2022[206]. If enactment, this reform would provide the first federal law that regulates the use of algorithms. Unlike the federal privacy regulation, there is no federal presumption of pre-emption within the AAA and states would therefore be free to

[205] Explanatory Memorandum at para 3.5
[206] Office of U.S. Senator Ron Wyden (2022a). Algorithmic Accountability Act (AAA) (2022). *117th Congress 2D Session*. https://doi.org/10.1016/S0140-6736(02)37657-8

impose higher standards on the use of AI in automated decisions if they deem appropriate[207]. In addition to this, there has been a significant increase in legislative activity within the individual states in the enactment of AI specific regulatory tools. Many of these emerging state laws are specific to the regulation of AI in an employment context and as such feature significant utility in protecting workers' rights from potential harms arising in context of automated and algorithmic decision-making in the workplace.

II.A. Algorithmic Accountability Act

Unlike the AIA, the AAA is framed for the regulation of automated decision-making rather than AI technologies in general and constitutes the first federal attempt to create a comprehensive AI regulatory framework that applies across industries. This is beneficial insofar that it reinforces the often misunderstood fact that automated decisions are a form of AI technologies and therefore synonymous to many but not necessarily all uses of AI. Instead of focusing on the regulation of high risk AI as the AIA does, the AAA focuses on 'augmented critical decision processes'[208] and requires large companies to audit these processes. The AAA does however contain additional requirements for critical decisions as well[209]. The scope of the AAA is narrower than the AIA since it only applies to "*covered entities*"[210]. The AAA also only applies to large companies or alternatively as an exception for the use of automated decision processes in smaller companies, who act as the de facto suppliers of the AI to the larger companies. The proposed act would require the Federal Trade Commission to design regulations for covered entities to perform impact assessments on any deployed augmented critical decision process or automated decision-making software where used by a covered

[207] § 11 AAA
[208] § 2(1)-(2) AAA
[209] § 2(8) AAA
[210] § 2(7) AAA

entity in the deployment of the augmented critical decision process[211]. Depending on the result of the assessment, covered entities would subsequently eliminate or mitigate any negative impact with legal or similarly legal effects on a consumer[212]. Different to the AIA where non-compliance would result in the issuance of administrative fines, violation of the AAA will amount to an unfair or deceptive act or practice[213]. The proposed act also gives the FTC the discretion to initiate further compliance measures if appropriate and the attorney general, or any other authorized government officer, the power to initiate a civil action[214].

II.B. Emerging State Laws

Several states have additionally enacted specific AI laws for particular uses of the technology in automated decision-making and profiling. In the context of hiring, Illinois passed the Artificial Intelligence Video Interview Act (AIVIA) 2020 which lays out notice, consent and explanation requirements for the use of AI in video and job interviews[215]. Other states, such as Maryland, have gone even further and banned the use of facial recognition software in job interviews unless the candidate provides a waiver[216]. New York has recently passed the Automated Employment Decision Tool Law (AEDT) *'*[217] which will impose a duty on employers to carry out bias audits of their automated employment tools prior to using such for hiring and firing decisions. However, unlike the earlier versions of the bill, the legislation in its current state does not impose the same regulatory threshold on other employment contexts and is therefore too narrow in its scope of enforcement[218]. They also will have to provide those

[211] § 3(b)(1)(A) AAA
[212] § 3(b)(1)(H) AA.
[213] § 9(a)(1) AAA
[214] § 9(a)(2)(D) AAA
[215] Artificial Intelligence Video Interview Act (2020)
[216] §3-717 Maryland Labor and Employment Code (2020)
[217] Local Law Int. No. 1894-A
[218] Matt Scherer and Ridhi Shetty, 'NY City Council Rams Through Once-Promising but Deeply Flawed Bill on AI Hiring Tools.' (2021) *Center for Democracy & Technology* NY City Council Rams Through Once-Promising but Deeply Flawed Bill on AI Hiring Tools – Center for Democracy and Technology (cdt.org) at paras 10-12

subject to the automated decision with detailed notice about the use of these tools as well as provide opportunity for an alternative selection process if so requested.

In addition the Stop Discrimination by Algorithm Act (SDAA) 2012 was recently introduced by the Attorney General of the District of Columbia. This act, if successful would make the use of AI discrimination unlawful both within and beyond the workplace by providing private individuals with rights of actions to tackle discrimination issues and requiring companies to give workers who are subject to automated decision-making both sufficient notice and explanations of such[219]. Likewise, the state of California is currently considering a Workplace Technology Accountability Act[220] which would impose limits on the use of electronic monitoring and automated decision-making of employees to specific times of day, geographical areas as well as for specific uses. The act would also provide workers with specific rights to challenge these activities, such as by providing information rights, corrective action rights and review powers, as well as impose obligations on employers to conduct algorithmic impact assessments to ensure the proper function of AI in the workplace.

Chapter 6: Conclusion

What is then learnt from the comparative regulation of AI in the workplace within both jurisdictions? Anti-discrimination and equality laws are regarded as the main device that police whether the use of algorithmic and automated decision-making can be regarded to be discriminatory or differential in treatment. Under these laws, AI must not lead to unfair differentiation of workers in the workplace. However, their practical effect on regulating AI is often limited since these emerging technologies are cultivating new, and unusual forms of

[219] Stop Discrimination by Algorithms Act of 2021, Office of the Attorney General of the District of Columbia
[220] Assembly Bill 1651 (AB 1651)

disparity such as intersectional, proxy-based and even novel forms of discrimination. These types of discrimination do not easily fit within the anthropocentric framework of the law that seeks to identify causation, intuition and in some cases even intention in order to find liability. This does not work with algorithms that are regression-based, iterative, and correlative in their computation process. Both systems therefore struggle to class algorithmic discrimination under the law of direct discrimination and disparate impact. In consequence, most cases of algorithmic discrimination will be subject to defenses that may exempt employers from liability. Within this anti-discrimination framework, US law is very willing to demonstrate a large amount of leniency towards the business and commercial incentives behind the use of algorithmic and automated decision-making. Although EU law is less included to do so, it is nonetheless not akin to a system of absolute or strict liability, with the effect that the provision of equality and anti-discrimination principles will be established on a case-by-case basis in many contexts of algorithmic and automated decision-making.

The alternative method of regulating algorithmic and automated decision-making in the workplace may be derived from data and privacy regulations. These laws are less concerned with the differential treatment of individuals who are subject to algorithmic and automated decision-making and more concerned with the transparency and accountability of the process and decision itself. Although data protection and privacy laws are more technologically inclined than anti-discrimination and equality laws they nonetheless are often limited and do not address the issues arising in relation to the need to obtain accessible, understandable and explainable decision-making processes in the workplace. Indeed, it is often possible to challenge algorithmic and automated processes more easily under these laws, but the results may nonetheless still not aid workers where the result is unintelligible or inaccessible to them due to the sheer complexity, and volume of data.

The final route of obtaining legal redress is therefore the current emergence of targeted and AI specific regulations that are being introduced in a piecemeal fashion in the EU and the US. These targeted regulations recognize and embrace the novelty of AI technologies and the unique challenges raised in the algorithmic and automated decision-making processes. Although similar in their respective objectives, the legal design of the lawmakers differs significantly in their approaches. The European AIA is based on broad problem-orientated principles that are modelled on a risk-based framework, which therefore elides many of the labor-specific issues raised by AI decision-making. In contrast, the piecemeal fashion of the US has resulted in many different legal solutions that are more sector specific and solution orientated. These regulations respond to AI by imposing audit requirements, impact assessments and sometimes even prohibitions where such is inherently deemed as risky. However, regulations are often myopically focused on particular solutions that they do not address the issue as a whole. As such, many states only regulate isolated practices of automated decision-making in the workplace such as the uses of AI in video-interviews or recruitment decisions. This fails to address the multifaceted operation and versatility of application of AI technologies.

The general law of anti-discrimination and data privacy may therefore serve as a fall back regime for workers in both jurisdictions who will fall outside of the targeted AI regulations. Workers seeking redress from discriminatory algorithms will be more protected in the EU than the US. In contrast, workers seeking protection relief to protect their privacy may, depending on the federal state they are in, be more successful in the US. This would be particularly true for workers who are subject to the CPRA where they would not only obtain protection against 'solely automated' but all automated decisions, irrespective of the effects of the AI harm, under the current legislative draft of the act. This not only imposes a stricter threshold of liability but potentially one that imposes a duty to provide for human decision-making where a workers chooses to opt-out of the automated process.

In so doing, both legal systems react with a constellation of laws to the emergence of AI technologies and embrace the impact of AI in the workplace context through a multi-sourced solution. Whether this will be sufficient to proact to new AI technologies is another question that will depend on the extent to which the AI legislative reforms rely on the subsidiary legal regimes that protect workers against AI through anti-discrimination and data protection laws. This will be a normative value judgement for lawyers, policy-makers and regulators to determine whether it is enough for the law to protect the modern workplace against the novelties of AI or whether these should adopt a more powerful and invasive regulatory framework as an exhaustive resolution to prepare the future workplace for these impeding technological developments.