WOMEN AT THE TABLE

# ARTIFICIAL INTELLIGENCE RECRUITMENT: Digital Dream or Dystopia of Bias?

WOMEN AT THE TABLE

Women At the Table is global civil society organization based in Geneva. It's the first organization to focus on systems change by helping feminists gain influence in sectors that have key structural impact: economy, democracy and governance, technology and sustainability.

Further information about Women At the Table can be found at **www.womenatthetable.net**

< 1 >

# Acknowledgments

The legal research and analyses for this report were conducted by Women At The Table. The principal authors are Caitlin Kraft-Buchman and Renée Arian.

In addition, attorneys from Skadden, Arps, Slate, Meagher & Flom LLP provided invaluable research and support in the preparation of the report, as follows: Kathlene Burke, Nicolas Etcheparre, James Anderson, David Herlihy, Eve-Christie Vermynck, Damian Babic, Jo Jimenez, Eleanor Williams, William Hardaway, Carine Louyot and Kimberly Franko Lower.

< 2 >

# Foreword

2020, the year COVID-19 caused global economic and social disruption, was meant to be a massive year of reflection and commemoration, even celebration, of women's rights.

2020 would be the marker of 25 years since the adoption of the Beijing Platform for Action, the visionary articulation of a path to gender equality drafted at an historic global gathering of women, by women, for women. 2020 would mark 5 years since every nation in the world, developed and developing, adopted the Sustainable Development Goals ("SDGs"), including the stand alone and transversal SDG #5 focused solely on gender equality. 2020 would mark the 10 years remaining to achieve the SDGs by 2030. 2020 would even mark the 40 year anniversary of the Convention to End All Forms of Discrimination Against Women ("CEDAW") often described as an international bill of rights for women, and 10 years since the landmark UN Security Council Resolution 1325 on Women, Peace and Security stressed the importance of women's equal participation and full involvement in all efforts for the maintenance and promotion of peace and security.

2020 was to be a critical year for the unfinished business of women's rights, because no one country has yet achieved gender equality, and the visionary aspirations of the Beijing Platform are mostly unfulfilled. 40 years after CEDAW, 25 years after Beijing, 10 years after UN SCR 1325, 5 years after the world agreed on the SDGs, we face most all of the same challenges in achieving equal rights, participation and influence for women. These equal rights and participation are increasingly fundamental to robust, resilient 21st century institutions, and prosperity, and a thriving global economy.
Instead, alarm bells have rung with concern that "decades of limited and fragile progress on gender equality and women's rights" have been reversed by the COVID-19 pandemic[1].

We face newly significant global challenges exposed by the COVID-19 pandemic. The Fourth Industrial Revolution, Artificial Intelligence and

**1** UN News, Generations of progress for women and girls could be lost to COVID pandemic, UN chief warns, UN News, Aug. 31, 2020, https://news.un.org/en/story/2020/08/1071372

Automated Decision-Making in machine learning offer new opportunities for achieving gender equality. This is good news. However, left unchecked, unaccountable and uncorrected they also present profoundly negative consequences for the ability of women to achieve their full participation and rights in the global workforce because of the historical bias hardwired into their data and larger systems. Predictive analytics, algorithms and other forms of AI offer hope for the perennial challenges of talent acquisition, seemingly producing higher dividends, superiority in candidate quality and a presumed reduction of discrimination in hiring.

Consequently, with remote working becoming standard practice, more companies and human resource management will lean on virtual AI tools to acquire talent—ranging from algorithms that produce ads targeted to 'ideal' candidates; programs that screen resumés, psychometric interview tests and gamified assessments; digital interviews that interpret facial expression; analytics dashboards that support final selection decisions; and observation and ranking of employee mood before and after client calls. AI is and will continue to be ubiquitous; impacting every aspect of the employment relationship everywhere, unseen—bringing us to this critical turning point in time.

Automation in recruitment is highly likely to reproduce and deepen the traditional biases reflected in existing data—data that is not only biased but incomplete because historically it has not, in general, included women. Automated recruitment /employment systems are being deployed, at a never before seen scale, and the incomplete data used to train these systems often causes them to infer that women are invisible]. Machine learning takes the missing information in the incomplete data—the invisible women—and makes the invisibility explicit in the code, mirroring the information received from the analog world and embedding it as permanent digital bias.

The exclusion of women from data gathering is far from novel. In fact, it is well documented from 20th century drug trials, international standards and global trading rules, to 21st century automated decision-making systems the default of a "standardized male" is used to form the physical framework and infrastructure of how we live and work. This is a dangerous default. We cannot afford to underestimate the urgency to find structural solutions and change course. This is particularly essential regarding the world of formal employment, as women's ability to participate in and contribute fully to rebuilding a post COVID-19 world is critical to social harmony and prosperity. This paper follows Women at the Table's report The Deadly Data Gap: Gender

& Data. When referring to women here, as in all of our work, we are speaking of all intersections of women and girls[2].

It is important to note that women can serve as a proxy for all or any groups traditionally invisible and 'other' to the system—those traditionally both left behind and made invisible. While the focus of this report is on gender discrimination against women in all their 'intersectionality'—i.e. the interconnected nature of social categorizations such as race, class and gender as they apply to a given individual or group, regarded as giving overlapping and interdependent systems of discrimination and disadvantage—the discussion is equally and powerfully applicable to other forms of discrimination, most notably racial discrimination[3].

Business as usual is no longer an option. In this time of upheaval in a world rocked by global pandemic, roiled by issues of race, gender and systemic bias, where inequities in our systems threaten our democracies, we also have an opportunity:  to create an inclusive digital landscape that advances the values of equality we have long espoused. We can forge a workplace culture where we all thrive. If done right, we can use technology as an aid to our ambitions of achieving equality. We have a moment, if we seize it now, to establish new norms with new technology that brings more equality—instead of enshrining inequality and systemic bias into the systems of our future.

**2** The Global Research Council defines intersectionality as the interconnected nature of social categorizations such as race, class and gender as they apply to a given individual or group, regarded as giving overlapping and interdependent systems of discrimination and disadvantage. Global Research Council, Supporting Women in Research: Policies, Programs and Initiatives Undertaken by Public Research Funding Agencies, (2019), https://anr.fr/fileadmin/documents/2019/GRC_GWG_Case_studies_final.pdf.

**3** Id

## < 2.1 > Landscape - Artificial Intelligence in Recruitment and Employment

The Council of Europe defines AI as a "set of sciences, theories and techniques whose purpose is to reproduce by a machine the cognitive abilities of a human being.[4]" The full achievement of this purpose – Artificial General Intelligence (AGI) or machines capable of independent thought and application of specialist learning to new contexts – is yet to be realized.
The AI that exists today is all Narrow AI – machines that carry out a specific task such as playing a game[5] or assessing candidates' performance at interview[6]. This Narrow AI ranges from simple algorithms to pattern-recognition systems to advanced applications such as Artificial Neural Networks that constantly and unpredictably adapt their analytical processes in pursuit of specific outcomes.

AI permeates modern recruitment from web crawlers to identify and attract favored candidates through applicant tracking systems, resumé content appraisers, gamified and classic assessments, automated interviews and interview analysis and candidate appraisal systems. A recent report estimated that 99% of Fortune 500 companies currently use Applicant Tracking Systems of some kind in their hiring process[7].

AI is expected to replace 16% of HR jobs within the next ten years[8], which

**4** Counsel of Europe, https://www.coe.int/en/web/artificial-intelligence/glossary

**5** See, e.g., DeepMind, https://deepmind.com/research/case-studies/alphago-the-story-so-far (detailing the story of AlphaGo, the first computer program to defeat a professional human Go player).

**6** See, e.g., HireVue, https://www.hirevue.com (a virtual interview assessment provider). See, also, Drew Harwell, A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job, Wash. Post, Nov. 6, 2019, https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job; David R. Fineman, People analytics: Recalculating the route: 2017 Global Human Capital Trends, https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2017/people-analytics-in-hr.html.

**7** Linda Qu, 99% of Fortune 500 Companies use Applicant Tracking Systems (ATS), Jobscan Blog (Nov. 7, 2019), https://www.jobscan.co/blog/99-percent-fortune-500-ats/.

**8** Forbes Coaches Council, 10 Ways Artificial Intelligence Will Change Recruitment Practices, Forbes (Aug. 10, 2018, 7:00 AM), https://www.forbes.com/sites/forbescoachescouncil/2018/08/10/10-ways-artificial-intelligence-will-change-recruitment-practices/#68b43cc13a2c.

means that corporate dependency on these software and processes will only continue to grow. Notably, men have a slightly more positive view of AI at work than women—32% of men are optimistic about AI at work compared to 23% of women[9]. The 'blended workforce'—humans and bots working side by side together[10]—is unlikely to diminish as a trend as 50% of employees in 2019 used some form of AI at work, up from 32% in 2018[11].

Human recruiting (without the use of AI) is by no means free from discrimination and bias[12]. Historical cases and studies have shown that human bias has led to discrimination and the law has developed in an effort to prevent discrimination and protect individuals from bias[13]. Many of these tools simply do not translate neatly to preventing bias in AI. For example, you cannot cross-examine AI in the same way that you can a hiring manager and particularly when dealing more complex versions of AI, it may be impossible[14]. Well-documented examples of AI systems reinforcing or perpetuating bias and gender discrimination in the workforce[15] are compounded by the historic underrepresentation of women in decision-

**9** Press Release, Oracle, New Study: 64% of People Trust a Robot More Than Their Manager (Oct. 15, 2019), https://www.oracle.com/corporate/pressrelease/robots-at-work-101519.html

**10** Jeanne Meister, Top 10 HR Trends That Matter Most in the 2020 Workplace, Forbes (Jan. 15, 2020 11:07 AM), https://www.forbes.com/sites/jeannemeister/2020/01/15/top-10-hr-trends-that-matter-most-in-the-2020-workplace/#320fe6a37dfc.

**11** Id.

**12** Devah Pager & Hana Shepherd, The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets, 34 ANN. REV. SOC. 181, 182 (2008).

**13** See, e.g., United Steelworkers of America v. Weber, 443 U.S. 193, 202-03 (1979).

**14** Pager & Shepherd supra note 12.

**15** See Ian Thomsen, Your Gender and Race Might Be Determining Which Facebook Ads You See, News@Northeastern (Apr. 11, 2019), https://news.northeastern.edu/2019/04/11/a-northeastern-university-team-tested-facebooks-algorithm-and-found-its-delivery-of-advertisements-is-skewed-by-race-and-gender/; Michael Rovatsos, Brent Mittelstadt & Ansgar Koene, United Kingdom Center for Data Ethics and Innovation, Landscape Summary: Bias In Algorithmic Decision-Making 19-22 (2019), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf. See also Caitlin Kraft-Buchman & Renée Arian, women@thetable, Triple A: Affirmative Action for Algorithms - Artificial Intelligence, Automated Decision-Making & Gender 3-7 (2019), available at https://uploads.strikinglycdn.com/files/6b4e02b8-5405-4087-927c-221a6b0608c8/WeShapeOurTools%20v11.pdf (citing several examples of gender bias in AI).

making positions[16] coupled with the reality that gender bias is pervasive in the underlying incomplete historical data[17].

As AI is ramped up at all stages of the employment relationship, it is important we develop new legal tools to prevent bias from creeping into the system, from focusing on the data that is being used to how the AI system ultimately provides a recommendation or makes a decision.

AI can be employed at all stages of the employment process, but is more commonly used in talent acquisition or the "hiring funnel": sourcing, screening; interviewing; and selection[18].

**16** Grant Thornton, Women in Business: Building a Blueprint for Action 2, 5-6 (Mar. 2019), https://www.grantthornton.global/globalassets/global-insights---do-not-edit/2019/women-in-business/gtil-wib-report_grant-thornton-spreads-low-res.pdf (noting that in 2019, 29% of senior management roles were held by women—the highest number ever on record. 43% of human resources directors are women compared to 17% of sales directors and 16% of chief information officers ); Rebecca Cassells & Alan Duncan, Bankwest Curtin Economics Centre, Gender Equity Insights 2019: Breaking Through the Glass Ceiling, 8, BCEC | WGEA Gender Equity Series 4 (Mar. 2019), available at https://bcec.edu.au/assets/2019/02/BCEC-WGEA-Gender-Equity-Insights-2019-Report.pdf (arguing that if the current growth patterns continue we will have to wait until 2100 for equal shares of female CEOs); Kraft-Buchman & Arian, supra note 15, at 17 (noting the situation is similarly replicated in the technology sector and in particular in AI and ADM leadership where women have remained largely under-represented and / or excluded).

**17** See Kraft-Buchman & Arian, supra note 15, at 3-7; Thomsen, supra note 15; Rovatsos, Mittelstadt & Koene, supra note 15, at 19-22 (noting that from 20th century drug trials, the design safety features in cars, medical treatments, the work equipment we wear, to name a few examples, are based on data that uses the default of a "standardized male."). See, also, Caitlin Kraft-Buchman & Renée Arian, women@thetable, The Deadly Data Gap: Gender & Data 3-8 (2019), available at https://uploads.strikinglycdn.com/files/cb0104a3-7f50-49c7-858e-827169cceabd/FINAL%20UPLOAD%20The%20Deadly%20Data%20Gap%20Gender%20&%20Data.pdf; W. Later et al., Is the 1975 Reference Man Still a Suitable Reference?, Eur. J. Clin. Nutr. 64(10), 1035-42 (2010), https://www.nature.com/articles/ejcn2010125.pdf; Mary Olson, Atomic radiation is more harmful to women, Wise (Nov. 11. 2011), https://www.wiseinternational.org/nuclear-monitor/736/atomic-radiation-more-harmful-women; Mary Olson, Females Exposed to Nuclear Radiation Are Far Likelier Than Males to Suffer Harm, PassBlue (July 5, 2017), https://www.passblue.com/2017/07/05/females-exposed-to-nuclear-radiation-are-far-likelier-than-males-to-suffer-harm/.

**18** See, e.g., Isaak, https://www.statustoday.com/ (last visited Nov. 2, 2020) (showcasing an AI employment product that analyses emails and accesses work files to show bosses how collaborative workers are and whether they are "influencers" or "change-makers"); ENGAGE Talent, https://www.engagetalent.com/ (last visited Nov. 2, 2020) (showcasing an AI product that can be used to analyze the workforce, engage individual employees in career coaching and 'predict' the likelihood that a candidate will change jobs).

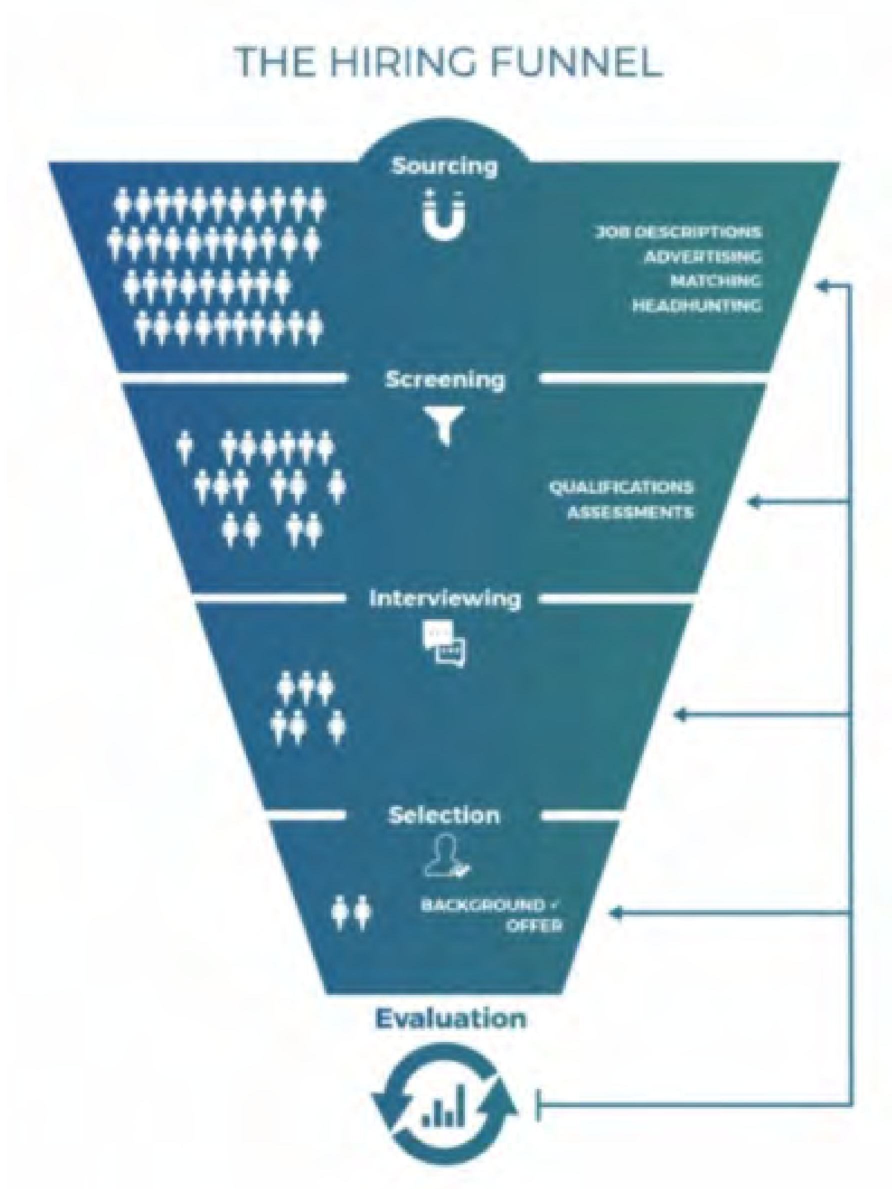


Image Source: Julius Schulte, AI-Assisted Recruitment Is Biased. Here's How To Make It More Fair, World Economic Forum (May 9, 2019), https://www.weforum.org/agenda/2019/05/ai-assisted-recruitment-is-biased-heres-how-to-beat-it.

## < 2.1.1 > Sourcing

The first step of the hiring funnel is sourcing, with its aim to generate a strong pool of applicants, accomplished via targeted job description, advertisement or active headhunting. This first outreach creates a critical first pool of applicants, widening the circle of possibilities or keeping the pool with the traditional same depth and dimension. Algorithms and AI systems are used to advertise openings, by optimizing job ads as well as their wording[19], notify potential applicants about appealing positions, tailor adverts to make open positions more attractive to specific profiles of candidates and identify prospective employees for recruiters[20] (e.g. identification programs that trawl CVs or websites for "ideal" candidates)[21].

Bias may be present in how job advertisements target potential employees. Personalized job boards aim to automatically learn recruiters' preferences and use those predictions to solicit similar applicants. This type of recommendation system is purposefully built to find and replicate patterns in user behavior. If the algorithm sees a pattern that recruiters interact more with white males, it makes the pattern a rule (an infamous example being the algorithm programmed to favor prospective employees named Jared with additional points if they played lacrosse in high school)[22]. The algorithm matched these characteristics to indicate positive job performance and then reinforced the pattern. This sort of adverse impact can happen without explicit instruction, and worse, without rigorous model testing can and frequently does go undetected[23]. A highly effective barrier to hiring certain

**19** See, e.g., Textio, https://textio.com/ (last visited July 5, 2020) (showcasing a platform that tailors job adverts using augmented writing so that they attract certain candidates).

**20** See Aurélien Rayer, Exploring Collaborative Filtering For Job Recommendations, Welcome to the Jungle (Feb. 1, 2019), https://www.welcometothejungle.com/en/articles/collaborative-filtering-job-recommendations (noting collaborative filtering is used to capture user preferences for both job seekers and recruiters).

**21** Forbes Coaches Council, supra note 8. See, e.g., Pymetrics, https://www.pymetrics.com/employers (last visited July 5, 2020) (showcasing a program where candidates play games and are then matched with opportunities, based on how existing employees have played the same games).

**22** See Kraft-Buchman & Arian, supra note 15, at 3-7; Thomsen, supra note 15; Rovatsos, Mittelstadt & Koene, supra note 15, at 19-22.

**23** Miranda Bogen, All The Ways Hiring Algorithms Can Introduce Bias, Harv. Bus. Rev. (May candidates can be achieved simply by not informing them of a job opportunity. Additionally, while not preventing an individual from applying for a position, web crawlers used in the sourcing process scan information from publicly available online sources to try match candidates to job descriptions. This screening process can reinforce strong biases against those who were not initially matched by the web crawlers that are not easy to overcome.

### < 2.1.2 > Screening

The second step of the hiring funnel is screening where a candidate's application is reviewed[24]. Keyword-driven sifting algorithms determine where and which categories candidates fit. This information could include job title, years of experience, languages, university degrees and countries where an individual has worked. Algorithms are used to reject candidates that do not conform to the parameters of the algorithm[25]. Other tools go further to make predictions based on past screening decisions. There are also processing programs that take basic details from job candidates, organize interviews and generally remove humans from the candidate sorting process[26].

Each of these processes has the potential for bias, which could arise from the use of the underlying data or the instructions programmed into the algorithm. For example, if the gender distribution of the training data was strongly imbalanced, this may be replicated by the algorithm, even if gender is not included in the information provided in the application documents. Bad data quality or small, non-diverse data sets could also result in training

6, 2019), https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias.

**24** Julius Schulte, AI-Assisted Recruitment Is Biased. Here's How To Make It More Fair, World Econ. F. (May 9, 2019), https://www.weforum.org/agenda/2019/05/ai-assisted-recruitment-is-biased-heres-how-to-beat-it.

**25** See Nuri Purswani, Word2Vec For Talent Acquisition, Medium.com (Sept. 13, 2018) https://medium.com/@nuripurswani/word2vec-for-talent-acquisition-ab20a23e01d8, (citing examples of software that uses learning vector representations of words called "word embeddings" and machine learning algorithms to screen resumes for keywords in context and to create relative rankings between the different candidates); see also Schulte, supra note 24.

**26** See Kumba Sennaar, Machine Learning For Recruiting And Hiring—6 Current Applications, Emerj (May 20, 2019), https://emerj.com/ai-sector-overviews/machine-learning-for-recruiting-and-hiring (citing examples of software that follows the general format of a chatbot and engages with potential candidates, handling tasks such as interview scheduling and responding to general inquiries).

data biases. Using biased data inputs and past hiring decisions to train the algorithm to evaluate who is most likely to be the "right" applicant could result in the algorithm inadvertently perfectly replicating and perpetuating the same bias that were present before the use of the AI recruiting tool[27].

Finally, resumés may be rejected merely because the algorithm was trained to read specific formats of documents and cannot interpret the data presented in a different 'non-conforming format'[28].

### < 2.1.3 > Interviewing

If an individual has made it through the screening process they may be invited to an interview where different algorithms may be used to support a hiring decision[29]. AI has introduced a new breed of psychometric tests during the interview (and also screening) stage of the employment selection process – gamified assessments[30]. These tools focus on enhancing candidate experience and apply game like features such as real-time feedback and interactive and immersive scenarios. The individual's choices and behaviors are matched by computer-generated algorithms to identify suitability for a given rol[31]. Unlike established psychometric tests, gamified assessments have yet to demonstrate robust longitudinal reliability, relevance, validity and

**27** See Schulte, supra note 24.

**28** Id. (citing examples in: Japan where there is a common resumé template (Rirekisho) used by all job applicants; China, where applicants list their work experience in the reverse chronology; the US, where resumés are usually one page with no photo; and in Europe, where resumés can be between two or three pages with a photo at the very top).

**29** Id.

**30** Tomas Chamorro-Premuzic, Dave Winsborough, Ryne A. Sherman & Robert Hogan, New Talent Signals: Shiny New Objects or a Brave New World?, 9 Indus. & Org. Psychology 621, (2016) available at https://pdfs.semanticscholar.org/83ef/b5454fd547d0284c89b2cbf0940ac14459b2.pdf.

**31** See Josh Bersin & Tomas Chamorro-Premuzic, New Ways To Gauge Talent And Potential, MIT Sloan Management Rev. (Nov. 16, 2018), https://sloanreview.mit.edu/article/new-ways-to-gauge-talent-and-potential; Melissa Hart, Subjective Decision making and Unconscious Discrimination, 56 Ala. L. Rev. 741 (2005), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=788066 (describing HireVue's MindX, which uses gamified cognitive-ability tests by asking users to play games; Pymetrics, which applies classic psychological tests in a game-like way; Arctic Shores, which puts candidates through a series of arcade games to correlates their choices to standard personality traits and competencies and Red Bull's Wingfinder, which attracts candidates through its social media channels and provides an extensive report on their strengths and weaknesses).

lack of bias[32]. In general, there is a worrying lack of independent research into these assessments and their potential for encapsulating gender and other biases[33].

Also used are digital interviews with algorithms used to flag and interpret facial expressions, tone of voice, language choice, speed, focus and emotions such as anxiety and excitement (amongst other interpretations). These AI tools use data-driven sorting and ranking to replace human observations and intuitive inference to identify and interpret "talent" signals[34].

If the training data used to evaluate video interviews (conducted prior to any in-person interaction) has not been thoroughly vetted against categories such as gender, age or religion, the same biases that were present before the use of the AI recruiting tool could be replicated.

Additionally, machine-learning software trained on these datasets don't just mirror biases, they can amplify them. For example, a major U.S. technology company claimed an accuracy rate of more than 97% for a face-recognition system they had designed. However, the data set was more than 77% male and more than 83% white[35]. In this case, the selection bias from data used to train the algorithm over-represented one population, while under-representing another.

---

**32** The British Psychological Society, Psychological testing: A test user's guide (2017), https://ptc.bps.org.uk/sites/ptc.bps.org.uk/files/guidance_documents/ptc02_test_users_guide_2017_web.pdf.

**33** Richard Landers, Developing a theory of gamified learning: Linking serious games and gamification of learning, 45(6) Simulation & Gaming 752 (2014), available at https://www.researchgate.net/publication/268632276_Developing_a_Theory_of_Gamified_Learning.

**34** Bersin & Chamorro-Premuzic, supra note 31.

**35** See Schulte, supra note 24; see also United Nations Educational, Scientific and Cultural Organization (UNESCO), I'd Blush If I Could: Closing Gender Divides In Digital Skills Through Education, 19 (2019), https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1; Larry Hardesty, Study Finds Gender And Skin-Type Bias In Commercial Artificial-Intelligence Systems, MIT News (Feb. 11, 2018), http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212; David Braue, ACS Information Age, Recruiters Tap AI for Post-COVID Staff Hunt, ACS (May 5, 2020), https://ia.acs.org.au/article/2020/recruiters-tap-ai-for-post-covid-staff-hunt.html; Kyle Wiggers, AI Weekly: Coronavirus Spurs Adoption of AI-powered Candidate Recruitment and Screening Tools, VentureBeat (Mar. 13, 2020), https://venturebeat.com/2020/03/13/ai-weekly-coronavirus-spurs-adoption-of-ai-powered-candidate-recruitment-and-screening-tools/.

## < 2.1.4 > Selection

Algorithms and analytics dashboards are used through the hiring process and to support the employer's final selection decision and generate ad hoc job offers[36]. These AI systems often measure signals related to tenure, productivity or performance (or the absence of signals such as tardiness or disciplinary action) or predict which applicants will be "successful" on the job[37].

Candidate data mining is a fast growing phenomenon throughout the hiring funnel. Employers passively mine candidate data by analyzing a candidate's digital footprints. Employers investigate a candidate's interests, personality, abilities, reputation and level of authority on social media and networking websites, which is used to theoretically predict a candidate's "suitability" for a position[38].

## < 2.1.5 > The "un-biased" automated recruitment system

Ironically, automated recruitment and employment tools were, and in some circles still are, thought to be able to themselves erase long standing unconscious (or even conscious) discriminatory hiring practices; stand-alone magic bullets to promote diversity and inclusion at work. Sadly this is not the case[39].

---

**36** Javier Sánchez-Monedero, Lina Dencik & 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, in ACM, FAT 2020: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, arXiv (Jan. 17, 2020), https://arxiv.org/pdf/1910.06144.pdf (explaining that candidates are assessed based on the keywords, facial expressions and tones they use in video interviews and that after a video interview, you may get a face-to-face interview that could lead to an offer).

**37** See Bogen, supra note 23.

**38** See Bersin & Chamorro-Premuzic, supra note 31 (explaining that by giving recruiters a range of scores to help evaluate candidates, social media platforms provide tools that do this automatically for recruiters); see also Sennaar, supra note 26; Golnoosh Farnadi et al., Computational Personality Recognition in Social Media, CORE 109-142, https://core.ac.uk/download/pdf/55893069.pdf.

**39** Sánchez-Monedero et al., supra note 36; Pauline Kim & Sharion Scott, Discrimination in Online Employment Recruiting, St. Louis Univ. L.J., 63(1) (2018), available at https://pa-

Although research shows that removing gendered language from job adverts, anonymizing applicants, using work sample tests, comparative evaluation on candidates, and other strategies[40] are all important and powerful levelers when applied, automated hiring systems themselves (particularly without blending human in the loop or scientifically-derived methods with multidisciplinary research and awareness to back them up) may not by themselves eliminate bias. To illustrate how nuanced gender bias is even with blinded review, gendered outcomes can still arise under anonymous evaluation according to a recent study that shows innovative research by women is underfunded because of simple gender differences in writing style[41].

A number of automated hiring systems aim to tackle discriminatory hiring practices and unconscious bias by replacing potentially biased human decision-making with what is perceived to be a "neutral technical" process, such as anonymizing candidates[42]. However, there are a number of limitations with using automated hiring systems that begin with on-going and inherent problems in the data. Indeed sample bias, label bias and even outcome bias can also apply to automated hiring systems. For instance, as an example of sample bias, the reference data itself can become the vehicle for bias if it is exacted from a current or past roster of employees who are homogeneous and non-diverse.

This sample bias makes data-driven predictions for what is a "good" or appropriate "fit" for an organization a self-fulfilling circle of homogeneity. Automated hiring systems have a single-axis understanding of identification which fails to engage substantially intersectional forms of discrimination. These issues become further amplified when considering language and its structure can contain "recoverable and accurate imprints of our historic biases", which can be "problematic as towards race or gender"[43].

pers.ssrn.com/sol3/papers.cfm?abstract_id=3214898; See Schulte, supra note 24.

**40** See Iris Bohnet, What Works: Gender Equality By Design (Belknap Press of Harv. Univ. Press, 2016).

**41** See Julian Kolev et al., Is Blinded Review Enough? How Gendered Outcomes Arise Even Under Anonymous Evaluation, National Bureau of Economic Research Working Paper Series (2019), https://gap.hks.harvard.edu/blinded-review-enough-how-gendered-outcomes-arise-even-under-anonymous-evaluation.

**42** See Sánchez-Monedero et al., supra note 36.

**43** Aylin Caliskan, Joanna J. Bryson & Arvind Narayanan, Semantics Derived Automatically

On top of the data issues, there are a host of issues surrounding accountability and transparency that notably begin with a lack of publicly available information on how automated hiring systems function. Jobseekers may never know whether an automated hiring system has been used in a successful or unsuccessful hiring process—that is, if the jobseekers even become aware of a potential opening in the first instance.

This lack of transparency may also serve to widen the inequality divide creating a "knows" and "knows-not" gap of those who are able to understand and therefore attempt to game the system. Additionally, automated hiring systems are often developed in one jurisdiction and used in another, which not only creates potential legal issues regarding differing discrimination and equality laws but also involves the differing data protection rights (further discussed in section 2.2 of this paper)[44] as well as general privacy issues pertaining to candidates[45].

It is clear that AI in recruitment and employment is changing the future of work. Compounded with an ever increasing reliance on data, which can further perpetuate the bias cycle through sample bias, label bias and outcome bias, our future is currently being built on technologies and decision-making systems that may embed offline discrimination and biases into the online world and in our digital work futures unless we resolve these issues[46].

from Language Corpora Contain Human-Like Biases, 356 SCIENCE 183 (2017), https://science.sciencemag.org/content/356/6334/183.

**44** See Australian Human Rights Commission, Human Rights And Technology: Discussion Paper December 2019 (2019), available at https://tech.humanrights.gov.au/sites/default/files/2019-12/TechRights2019_DiscussionPaper.pdf.

**45** Ben Dattner, Tomas Chamorro-Premuzic, Richard Buchband & Lucinda Schettler, The Legal and Ethical Implications of Using AI in Hiring, Harv. Bus. Rev. (Apr. 25, 2019), https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring.

**46** Iris Bohnet, Real Fixes for Workplace Bias, Wall St. J. (Mar. 11, 2016), https://www.wsj.com/articles/real-fixes-for-workplace-bias-1457713338 ("A review of almost 1,000 studies on interventions aimed at reducing prejudice found that most programs weren't tested. For the few that were, including media campaigns and corporate-diversity training, the effects, wrote Elizabeth Levy Paluck of Princeton and Donald P. Green of Yale in the Annual Review of Psychology (2009), "'remain unknown.'").

< 3 >

# The Current Legal Framework

There are two main areas of law that aim to protect women against discrimination by AI systems used in the hiring process: employment law, that deals directly with the issue of discrimination, and data protection, that also provides protection to women against discrimination, in particular by protecting the way they are treated as data-subjects.

Therefore, focusing on the United States, United Kingdom and France (the **"In-scope Jurisdictions"**), the report looks at legal frameworks in the In-scope Jurisdictions that govern discrimination and data protection, as well as other frameworks from which inspiration might be drawn to draft regulation pertaining specifically to discrimination and AIs.

## < 3.1 > Employment Law

Gender discrimination in an employment context, whether perpetrated by a human or AI bias, is unlawful in the In-scope Jurisdictions. In each of these jurisdictions, a patchwork of laws prevent gender discrimination in the workplace. Little of the legislation directly addresses AI technology currently in the market since much of the legislation was drafted and introduced prior to the quotidian use of AI. The siloed approaches within the In-scope Jurisdictions have resulted in an accountability gap. As countries continue to develop legislative frameworks grounded in human rights, coherence and consistency between jurisdictions is needed.

## < 3.1.1 > United Kingdom

In the UK, gender is one of a number of protected characteristics under the Equality Act 2010[47]. It is unlawful in the UK to directly discriminate against an employee because of a protected characteristic. For example, rejecting female applicants' CVs purely on the basis of gender would be direct sex discrimination.

Indirect discrimination may also be unlawful. If an employer operates a Provision, Criterion or Practice (a "**PCP**") that on its face applies equally to all employees regardless of whether they have a protected characteristic or not, but in fact puts a certain group of employees at a disadvantage because of their protected characteristic, then this PCP may amount to unlawful indirect discrimination. Similarly, an employer penalizing CVs that contain periods of time off from work could amount to indirect discrimination because women have been more likely to have taken time off from work for care commitments, or an employer requiring that job applicants be of a certain height unless it could be proved it was necessary to the proper execution of the job[48].

Therefore, if the use of AI by an employer has the effect of disadvantaging female employees as a group compared to their male counterparts, then that could amount to indirect discrimination[49]. For example, AI built upon biased datasets or unconscious bias by advertising certain jobs to men only or penalizing CVs that show 'female characteristics' can put women at a particular disadvantage[50]. There is no need for the employer to have intended to cause discrimination: the test is, based on the particular facts of the matter, whether the use of AI did have a discriminatory effect.

**47** Equality Act 2010, c. 15, § (4) (U.K.) (listing "protected characteristics", which also include age, disability, gender reassignment, race, religion or belief, sexual orientation, marriage and civil partnership and sex).

**48** See CJEU Judgment: Police Rule on Height Limit May Discriminate Against Women, EUR-LEX, https://eur-lex.europa.eu/content/news/CJ_police_height.html (last visited July 5, 2020).

**49** Equality Act 2010, c. 15, § 19 (U.K.).



**50** See Kraft-Buchman & Arian, supra note 15, at 4.

If the particular provision, criterion or practice, in this instance the use of AI in recruitment, can be objectively justified, in order to defend an indirect discrimination claim, an employer must prove that the use of AI is a proportionate and effective means of achieving a legitimate aim[51]. By way of consequence, even if an employer unintentionally uses a discriminatory AI in their recruitment process, for example in order to find the best candidates for any open job position, the practice might have a legitimate aim but it could be determined not be an effective means of achieving that aim in every circumstance. If that were the case, the employer would be found to have indirectly discriminated against job candidates, regardless of the employer's intention.

### < 3.1.2 > France

France like all European Union countries, has implemented the Employment Equality Framework Directive (Council Directive 2000/78/EC), which underpins the European legal regime for sex discrimination at work. A number of other notable Directives relevant to equality and bias includes the Equal Opportunities and Equal Treatment Directive 2006/54/EC, the Gender Equality Directive 2004/113/EC. Human rights standards under the European Convention of Human Rights ("ECHR") are also relevant as they refer to protection against discrimination (article 14) on any ground in relation to rights and freedoms guaranteed by the ECHR. Therefore, the regime in France is not dissimilar to that in the UK.

Also, according to the French Constitution, the law must guarantee equal rights between men and women[52]. Article L. 1132-1 of the French Labor Code prohibits any type of discrimination in the course of employment based on certain protected characteristics, such as race and gender[53]. Violation of this statutory provision is a criminal offence, provided by articles 225-1 et seq of the French criminal Code[54], exposing the legal representative of the company to a maximum 3 year prison sentence and €45,000, and exposing

**51** Equality Act 2010, c. 15, § 19(2)(d) (U.K.).

**52** 1958 Const. (Fr) art. 1.

**53** Code du travail [C. trav.] [Labor Code] art. 1132-1 (Fr.).

**54** Code penal [C. pén.] [Criminal Code] art. 225-1 (Fr.).

the company to a €135,000 fine as well as civil claims[55].

French law also recognizes the existence of indirect discrimination, and similarly to the UK, any practice that might seem neutral in appearance but in practice discriminates against individuals based on protected characteristics would be considered discriminatory[56]. If so, the sanctions provided for discrimination by the French Labor Code and Criminal Code would apply[57]. This principle was applied for the first time with regards to gender discrimination in a decision rendered by the French Supreme Court for judicial matters, in which it found a company's refusal to register certain job positions, prominently occupied by women, to a specific retirement

**55** Code penal [C. pén.] [Criminal Code] art. 225-2 (Fr.) ("The discrimination defined in Articles 225-1 to 225-1-2, committed against a natural or legal person, is punishable by three years' imprisonment and a fine of 45,000 euros when it consists of:

1° Refusing to supply a good or service;

2° hindering the normal exercise of any economic activity whatsoever;

3° Refusing to hire, punish or dismiss a person;

4° To subject the supply of a good or service to a condition based on one of the elements referred to in Article 225-1 or provided for in Articles 225-1-1 or 225-1-2;

5° To make an offer of employment, an application for a training course or a period of training in a company subject to a condition based on one of the elements referred to in Article 225-1 or provided for in Articles 225-1-1 or 225-1-2;

6° To refuse to accept a person for one of the training courses referred to in 2° of Article L. 412-8 of the Social Security Code.

Where the discriminatory refusal provided for in 1° is committed in a place open to the public or for the purpose of prohibiting access thereto, the penalties are increased to five years' imprisonment and a fine of 75,000 euros.")

**56** Code du travail [C. trav.] [Labor Code] art. 1132-1 (Fr.)("No person may be excluded from a recruitment or appointment procedure or from access to an internship or training period in the company, no employee may be sanctioned, dismissed or be the subject of a discriminatory measure, whether directly or indirectly, as defined in Article 1 of Law 2008-496 of May 27, 2008 containing various provisions for adaptation to Community law in the area of the fight against discrimination, in particular with respect to compensation, as defined in Article L. 3221-3, profit-sharing or share distribution measures, training, reclassification, assignment, qualification, classification, professional promotion, transfer or renewal of contracts on the grounds of origin, sex, morals, sexual orientation, gender identity, age, marital status or pregnancy, genetic characteristics, or particular vulnerability resulting from the economic situation, apparent or known to the perpetrator, his or her membership or non-membership, real or supposed, to an ethnic group, nation or alleged race, his or her political opinions, his or her union or mutualist activities, his or her exercise of an elective mandate, his or her religious convictions, his or her physical appearance, his or her surname, his or her place of residence or his or her bank account, or because of his or her state of health, loss of autonomy or disability, his or her ability to express himself or herself in a language other than French.").

**57** Code penal [C. pén.] [Criminal Code art. 225-1 et seq (Fr.).

fund[58] was indirect discrimination, even if its goal was to ensure the financial stability of that retirement fund. Therefore, if AI uses relevant features that are tainted (as a result of biased data) with protected characteristics, including race and gender, in its decision-making process that would be considered discriminatory as such characteristics cannot be used by employers to justify a decision pertaining to a candidate or employee.

## < 3.1.3 > United States

### < 3.1.3.1 > Anti-discrimination law

Under US federal law, sex (including gender) is a protected characteristic under Title VII of the Civil Rights Acts of 1964 ("**Title VII**")[59]. Title VII prohibits direct discrimination against an applicant or employee on the basis of a protected characteristic[60]. Title VII also prohibits unintentional discrimination, or disparate impact discrimination[61]. Disparate impact discrimination occurs when an apparently neutral policy or practice unduly disadvantages individuals based on a protected characteristic[62]. To ensure compliance, the Civil Rights Acts of 1964 established the Equal Opportunity Commission ("**EEOC**") which in turn issued the Uniform Guidelines on Employment Selection Procedures in 1978, to set the standards for how employers can choose their employees[63].

To bring a disparate impact claim under Title VII the individual must show that the policy or practice has a disproportionately harmful effect based

**58** Cour de Cassation [Cass.], Supreme Court for Judicial Matters, soc. June 6, 2012, Bull. civ. V, No. 168.

**59** Civil Rights Act of 1964, 42 U.S.C. § 2000e et seq (1964); See, also, National Conference of State Legislatures, Sex and Gender Discrimination in the Workplace, https://www.ncsl.org/research/labor-and-employment/-gender-and-sex-discrimination.aspx (noting that many states and municipalities in the US have separate laws which also prohibit sex or gender based discrimination).

**60** Civil Rights Act of 1964, 42 U.S.C. § 2000e-2

**61** Id.

**62** Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 987 (1988).

**63** Civil Rights Act of 1964, 42 U.S.C. § 2000e-4; See, also, Employment Tests and Selection Procedures, U.S. Equal Emp't Opportunity Commission, https://www.eeoc.gov/laws/guidance/employment-tests-and-selection-procedures (last visited July 5, 2020).

on the protected characteristic (e.g. sex); which is usually done through statistical comparisons[64]. When using statistical comparison, the individual generally must show that the selection rate of the class of individuals with the protected characteristic is less than four-fifths (or 80%) of the selection rate for the most successful group[65]. If the individual is able to show this, then the burden of persuasion shifts to the employer. The employer must then show that the policy or practice is job related for the position in question, consistent with business necessity and no other alternative employment requirement suffices[66]. If the employer meets this standard, the individual can prevail by showing that an alternative employment practice has less disparate impact and also serves the employer's legitimate business interest[67]. That is, the standard does not prohibit disparate impact all together.

For example, if a business that used a screening algorithm can demonstrate that the algorithm has validity (meaning it can accurately predict a job related quality) this can be a defense against a claim of disparate impact. Thus, even if the screening algorithm did produce a disparate impact, it could be justified as serving a legitimate business objective if it is sufficiently accurate[68].

AI used in recruiting and employment decisions can still produce a disparate impact if it uses variables that are correlated with both the output variable the AI system is trying to predict and a variable for protected-class status, even if developers deliberately avoided using variables for protected classes to begin with[69]. Therefore, the legal risk of using AI arises more from the exposure to claims of disparate impact than in intentional discrimination.

**64** Watson, 487 U.S. 987.

**65** Civil Rights Act of 1964, 42 U.S.C. § 2000e-2(k).

**66** Id.

**67** Id.

**68** See Manish Raghavan & Solon Barocas, Report, Challenges For Mitigating Bias In Algorithmic Hiring, Brookings (Dec. 6, 2019), https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/.

**69** See Manish Raghavan & Solon Barocas, Report, Big Data's Disparate Impact 104 Cal. L. Rev. 671, 691 (2016).

### < 3.1.3.2 > AI specific regulation

U.S. law makers have already raised concerns of the potential risks of discrimination tied to the use of AI in recruitment, as demonstrated by a letter sent, in 2018, by three senators to various U.S. agencies, in particular the EEOC in which they raise concerns over the use of facial analysis software by employers during interviews, and request that the EEOC provide guidelines for employers eager to use such software[70].

In April 2019, a bill called the Algorithmic Accountability Act of 2019 was introduced in the House of Representatives, and an identical bill was introduced in the Senate, which would direct the Federal Trade Commission to require entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments, in particular for high-risk automated decision support algorithms[71]. "High-risk" automated decision support algorithms would include any algorithm that "**makes decisions, or facilitates human decision making, based on systematic and extensive evaluations of consumers, including attempts to analyze or predict sensitive aspects of their lives, such as their work performance.**"[72] The impact assessment must provide "**an assessment of the risks posed by the automated decision system to the privacy or security of personal information of consumers and the risks that the automated decision system may result in or contribute to inaccurate, unfair, biased, or discriminatory decisions impacting consumers.**"[73] At the time of writing, the bill has not passed and was referred to the House Committee on Energy and Commerce, which referred the bill to the Subcommittee on Consumer Protection and Commerce, and the identical bill in the Senate was referred to the Senate Commerce, Science, and Transportation Committee[74].

**70** Letter from Kamala D. Harris, U.S. Senator, et al. to Victoria A. Lipnic, Acting Chair, U.S. Equal Opportunity Emp't Comm'n, et al. (Sept. 17, 2018), available at https://www.hrpolicy.org/downloads/2018/Sen%20Harris%20Ltr%20to%20EEOC%20on%20Facial%20Recognition%20Software.pdf

**71** H.R. 2231, 116th Cong. (2019); S. 1108, 116th Cong. (2019).

**72** Id. § 2(7)(B) (emphasis added).

**73** Id. § 2(2)(C) (emphasis added).

**74** S.1108 - Algorithmic Accountability Act of 2019, Congress, https://www.congress.gov/bill/116th-congress/senate-bill/1108/committees (last visited July 5, 2020).

Certain States have also passed specific legislation aiming to regulate the use of AI in the recruitment process. For instance, on January 1, 2020, the Illinois' Artificial Intelligence Video Interview Act ("AIVI Act") took effect, which regulates employers' use of algorithms and other forms of AI to analyze video interviews[75]. Under this legislation, employers using AIs to analyze video interviews must:

– notify each applicant in writing before the interview that AI may be used to analyze the applicant's facial expressions and consider the applicant's fitness for the position;

– provide each applicant with an information sheet before the interview explaining how the AI works and what characteristics it uses to evaluate applicants; and

– obtain written consent from the applicant to be evaluated by the AI program[76].

The AIVI Act provides that an employer may not use AI to evaluate applicants who have not consented to the use of AI analysis[77]. Similarly, a new law will go into effect in Maryland on October 1, 2020, which prohibits employers from using certain facial recognition services for the purposes of creating a facial template in the interview unless an employee consents[78].

Illinois lawmakers also have been considering legislative solutions to prevent bias when employers use data analytics to make employment decisions. For example, two companion bills proposed in Illinois would prevent employers that use predictive analytics in making credit or hiring decisions from including information that correlates with the applicant's race or zip code with suitability for granting credit or employment[79]. HB 3415 was assigned to the Labor and Commerce Committee in March 2020[80]. HB 2991 has not

**75** Artificial Intelligence Video Interview Act, 820 ILCS 42/1-15 (2019)

**76** 820 ILCS 42/5.

**77** Id.

**78** H.B. 1202, 441st Gen. Assemb., Reg. Sess. (Md. 2020).

**79** H.B. 2991, 101st Gen. Assemb. (Ill. 2019); H.B. 3415, 101st Gen. Assemb. (Ill. 2019).

**80** Bill Status of HB3415, Ill. Gen. Assembly, http://www.ilga.gov/legislation/BillStatus.asp?DocTypeID=HB&DocNum=3415&GAID=15&SessionID=108&LegID=120007.

progressed beyond the Rules Committee[81]. The bills both define "predictive data analytics" as the use of automated machine learning algorithms for the purpose of statistically analyzing a person's behavior. If enacted, the laws would:

– Cover employers that both:
– use predictive data analytics to evaluate job applicants; and
– hire more than 50 Illinois residents in a calendar year.
– Require covered employers to devise procedures to ensure that the employers do not inadvertently consider information correlating with race or zip code when making hiring decisions[82].

Even more recently, on February 27, 2020, the New York City Council introduced a bill aiming to regulate the use of automated employment decision tools[83]. The bill proposes to prohibit the sale of such tools if they were not the subject of an audit for bias in the past year prior to sale, were not sold with a yearly bias audit service at no additional cost, and were not accompanied by a notice that the tool is subject to the provisions of the proposed bill. The bill also proposes a civil penalty of $500 for the first individual violation, with up to a $1,500 penalty for each subsequent violation. If passed into law, the bill would be enforceable starting January 1, 2022. Currently, the bill has been referred to the city's Committee on Technology[84].

## < 3.2 > Data Protection Laws

While data protection laws do provide some protection for individuals against discrimination, their essential weakness is that this is not their main purpose. Nonetheless, they do provide certain protections and can serve as a framework for strengthening legal frameworks to prevent discrimination or discriminatory impact.

**81** Bill Status of HB2991, Ill. Gen. Assembly, http://www.ilga.gov/legislation/BillStatus.asp?DocTypeID=HB&DocNum=2991&GAID=15&SessionID=108&LegID=119328. ]

**82** H.B. 2991, 101st Gen. Assemb. (Ill. 2019); H.B. 3415, 101st Gen. Assemb. (Ill. 2019).

**83** New York City, N.Y., A Local Law to amend the administrative code of the city of New York, in relation to the sale of automated employment decision tools, Int 1894-2020 (2020).

**84** Legislation, N.Y.C. City Council, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search (click Details tab) (last visited July 5, 2020).]

## < 3.2.1 > EU General Data Protection Regulation

The General Data Protection Regulation ("**GDPR**")[85] applies to the use of AI in recruitment and employment if such use involves the processing of personal data, such as names, dates of birth and previous work experience[86].

The GDPR contains general provisions relevant to the use of AI in recruitment and employment, regardless of whether the AI is used as part of a decision-making process in which there is no meaningful human involvement ("solely automated") or a decision-making process in which there is meaningful human involvement ("non-solely automated")[87]. The GDPR also contains separate provisions specific to solely automated decision-making.

### < 3.2.1.1 > General Provisions

The GDPR requires the processing of personal data to be lawful, fair and transparent[88]. It contains various lawful grounds for the processing of personal data[89]. When processing personal data for recruitment and employment, companies are likely to rely on the consent of individuals, the company's legitimate interests or the fact that the processing is necessary

**85** Commission Regulation 2016/679, 2016 O.J. (L 119) [hereinafter General Data Protection Regulation 2018].

**86** Who Does The Data Protection Law Apply To?, Eur. Commission, https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/application-regulation/who-does-data-protection-law-apply_en (last visited June 8, 2020) ("The GDPR applies to a company or entity which processes personal data as part of the activities of one of its branches established in the EU, regardless of where the data is processed; or a company established outside the EU and is offering goods/services (paid or for free) or is monitoring the behavior of individuals in the EU.").

**87** What Does The GDPR Say About Automated Decision-making and Profiling?, Information Comm'r's Office (u.k.), https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-does-the-gdpr-say-about-automated-decision-making-and-profiling/ (last visited June 8, 2020) ("A process might still be considered solely automated if a human inputs the data to be processed, and then the decision-making is carried out by an automated system. A process won't be considered solely automated if someone weighs up and interprets the result of an automated decision before applying it to the individual.").

**88** General Data Protection Regulation 2018, art. 5(1)(a).

**89** Id. art. 6(1).

for the performance of an employment contract. Companies wishing to rely on consent must be able to show that individuals understand what they are consenting to. This will at times be difficult given the technical nature of AI[90]. Alternatively, a company seeking to base the processing of personal data on its legitimate interests will have to assess whether the company's interests are overridden by the data subject's "interests or fundamental rights and freedoms.[91]" This balancing exercise, considering factors such as the risk of discrimination, will generally be documented in a "Legitimate Interest Assessment."

Regarding transparency, Article 12(1) of the GDPR requires individuals to be provided with "concise, transparent, intelligible and easily accessible information" about the processing of their personal data. This is usually done by way of a privacy notice. Articles 13 and 14 of the GDPR provide the specific requirements for the information to be provided and require individuals to be informed of, among other things, the purpose and legal basis for the processing of their personal data. In certain circumstances, discussed in the "Specific provisions" section below, this includes being informed of the use of AI.

As to fairness, the GDPR does not provide a great deal of clarity on what is meant by this term. However, it is noted in regulatory guidance related to the GDPR that profiling "may be unfair and create discrimination, for example by denying people access to employment opportunities.[92]"

The GDPR includes the overarching principle that companies must be accountable for their compliance and must be able to demonstrate compliance at any time if required to do so[93]. To help demonstrate this, companies are required to maintain a written record of their data processing activities[94]. Companies are also expected to have a data protection governance structure in which competent individuals in relevant positions have oversight of data

**90** Id. art. 30.

**91** Id. art. 6(1)(f).

**92** Guidelines on Automated Individual Decision-making and Profiling for the Purposes of Regulation 2016/679, Eur. Commission 10 (2017), available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (follow hyperlink) [hereinafter The EU Guidelines].

**93** General Data Protection Regulation 2018, art. 5(2)

**94** Id. art. 30

protection compliance and take responsibility for the implementation and monitoring of data protection policies, notices and procedures.

The principle of accountability is furthered by the requirement for a Data Protection Impact Assessment ("DPIA") when a type of processing is "likely to result in a high risk to the rights and freedoms of natural persons.[95]" Regulatory guidance on the GDPR further elaborates that "the reference to "the rights and freedoms" of data subjects primarily concerns the rights to data protection and privacy but may also involve other fundamental rights such as... prohibition of discrimination.[96]" In light of this, it is highly likely that a DPIA will be required where a company is using AI in its recruitment and employment process and should take potential discrimination into account. The GDPR seeks to avoid discrimination at an early stage through the codified notion of data protection by design and the requirement for a DPIA prior to processing that "is likely to result in a high risk to the rights and freedoms of natural persons."

A DPIA requires the following steps to be taken: (i) identify the need for a DPIA; (ii) describe the processing; (iii) consider consultation; (iv) assess necessity and proportionality; (v) identify and assess risks; (vi) identify measures to reduce risk; (vii) sign off and record outcomes; (viii) integrate outcomes into plan; and (ix) keep the DPIA under review[97]. As is made clear by step (ix), a DPIA requires ongoing review to ensure continued compliance with data protection and equality laws. Further, if the DPIA indicates that the processing would result in a high risk that cannot be mitigated, the local supervisory authority (such as the CNIL in France or the ICO in the UK) has to be consulted prior to the processing taking place[98].

**95** Id. art. 35(1).

**96** Data Protection Working Party, European Comm'n, Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679, at 6 (last revised and adopted Oct. 4, 2017), available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236 (follow hyperlink).

**97** Data Protection Impact Assessments, Information Comm'r's Office (U.K.), https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/ (last visited June 9, 2020).

**98** General Data Protection Regulation 2018, art. 36(1).

The DPIA should also facilitate early engagement (as a DPIA has to be carried out before the processing actually takes place) with all relevant departments, such as IT and HR within an organization. In theory this should help to ensure that any risk of discrimination is prevented or at least mitigated at an early stage by bringing together various complementary expertise.

Finally, the GDPR requires that data processing systems are designed to implement the data-protection principles. These include lawful, fair and transparent processing and accountability (the so-called "data protection by design" principle)[99]. This will mean that AI systems should have been developed with the data protection principles in mind from the outset and have the necessary safeguards in place.

### < 3.2.1.2 > Provisions specific to human in the loop vs. solely automated decision-making and profiling

The European Commission's 'Ethics Guidelines for Trustworthy AI' refer to various levels of human involvement in AI-informed decision-making, including human-in-the-loop, human-on-the-loop, and human-in-command[100]. Human-in-the-loop refers to capability for human intervention in every decision cycle of the system (noting the Guidelines recognize that in many cases this is neither possible nor desirable)[101]. Human-on-the-loop refers to human intervention during the design phase and monitoring of the system in operation. Human-in-command is the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. This includes decisions not to use AI, establishing levels of human discretion and giving the human decision maker the ability to override a decision[102]. The level of human involvement within automation has been shown to contribute to accountability within the automation system and ultimately in mitigating discrimination and bias[103].

**99** Id. art. 25

**100** Ethics Guidelines For Trustworthy AI, Eur. Commission (Apr. 8, 2019), https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

**101** Id.

**102** Id.

**103** See Australian Human Rights Commission, supra note 44, at 7.

Regulatory guidance has stated that to qualify as "human involvement" the involvement must be meaningful and not merely a token gesture[104]. While these guidelines are helpful in some respects, what exactly this entails is not clear but it has been suggested that factors such as automation bias (the tendency to agree with the outcome of the AI system) and the interpretability of the AI system are both factors that could prevent human involvement from being meaningful[105].

The GDPR prohibits decision-making that is solely automated which has legal or similarly significant effects for an individual[106]. This prohibition is subject to certain limited exceptions, for instance where the individual has explicitly consented to such processing of their personal data[107]. Solely automated decision-making is narrowly defined under the GDPR: there must be no, or very marginal, human influence in reaching the decision.

Where decision-making that is solely automated is allowed, the GDPR requires numerous safeguards to protect the rights, freedoms and legitimate interests of the individual. For instance, in certain circumstances, the individual will have the right to a review of the decision by someone who has the appropriate authority to overturn the decision[108]. In addition, individuals must be informed that the decision-making is solely automated, including "meaningful information about the logic involved, as well as the significance and envisaged consequences" of such processing[109]. As per regulatory guidance on the GDPR, the information provided must be sufficient for an individual to understand the reasons for a decision but need not entail "a complex explanation of the algorithms used or disclosure of the full algorithm.[110]"

**104** The EU Guidelines, supra note 92, at 21.

**105** See Reuben Binns & Valeria Gallo, Automated Decision-making: The Role of Meaningful Human Reviews, Computers & L., June 2019, at 29, 30-31.

**106** The EU Guidelines supra note 92, at 22 (noting decisions that "deny someone an employment opportunity or put them at a serious disadvantage" would be considered sufficiently significant to meet the legality threshold).

**107** General Data Protection Regulation 2018, art. 22(2).

**108** The EU Guidelines, supra note 92, at 27.

**109** Id. at 20.

**110** Id. at 25.

Unfortunately, in light of the protections required to carry out solely automated decision-making, and the relatively low barrier to avoid decision-making being classed as solely automated, it is unlikely that companies will fall within the remit of Article 22 of the GDPR—so it will rarely apply in practice.

### < 3.2.2 > United States data protection

The United States does not have a comprehensive federal law that regulates the collection and use of personal information, despite the Federal Trade Commission continually calling for flexible and technologically neutral privacy and security laws to be enacted[111]. Instead, the US Government has regulated certain sectors and types of sensitive information such as personal health, financial and children's information[112]. Additionally, many of the states and territories have laws seeking to safeguard data, disposal of data and privacy (amongst other things)[113]. California alone has more than 25 state data security and privacy laws including the most comprehensive privacy legislation in the US to date, the California Consumer Privacy Act of 2018 ("**CCPA**")[114], which came into effect on January 1, 2020 and mimics the GDPR in many ways[115].

Under the GDPR, data subjects have the right to not be subject to automated decision making, including profiling, which has legal or other significant effects on the data subject, subject to certain exceptions. This is not the

**111** Reforming The U.S. Approach to Data Protection and Privacy, Council on Foreign Rel. (Jan. 30, 2018), https://www.cfr.org/report/reforming-us-approach-data-protection; U.S. Fed. Trade Comm'n, Protecting Consumer Privacy in an Era of Rapid Change Recommendations for Business and Policymakers 12-13 (Mar. 2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

**112** See, e.g., Fair Credit Reporting Act (FCRA) 15 U.S.C. § 1681 (1992); Health Insurance Portability and Accountability Act (HIPPA), Pub. L. No. 104-191, 100 Stat. 2548 (codified as amended in scattered sections) (1996); Children's Online Privacy Protection Act (COPPA), 15 U.S.C. §§ 6501-6505 (1998). See, also FTC Website, https://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy-security.

**113** Congressional Research Service, Data Protection Law: An Overview, R45631 (Mar. 25, 2019), available at https://crsreports.congress.gov/product/pdf/R/R45631.

**114** California Consumer Privacy Act of 2018, Cal. Civ. Code §§ 1798.100 to 1798.199.

**115** Stuart D. Levi, Skadden's 2019 Insights: California Privacy Law: What Companies Should Do to Prepare in 2019 (Jan. 17, 2019).

case under the CCPA[116]. However, the November 2020 ballot includes a proposition to amend the California consumer privacy laws through the California Privacy Rights Act, which provides new data rights for consumers and places stricter obligations on companies[117]. Included in this proposal is to require transparency around automated decision-making and profiling, so that consumers can know how algorithms are evaluating them in the ways that affect job offers they see, the loans they are eligible for, and other decisions that affect their lives[118].

**116** Final Proposed California Consumer Privacy Act Regulations Submitted for Administrative Review, Privacy & Cybersecurity Update (Skadden, Arps, Slate, Meagher & Flom LLP New York, N.Y.), June 2020.

**117** Qualified Statewide Ballot Measures, Proposition 24, SECRETARY OF STATE OF CALIFORNIA, https://www.sos.ca.gov/elections/ballot-measures/qualified-ballot-measures (last visited September 15, 2020).

**118** Id. at 43.

< 4 >

# Roadmap to avoiding discrimination in AI and recruitment

Current legislation faces three major challenges with the advent of AIs and their use in recruitment and employment.

The first is scale, since a single but widely used discriminatory AI will have a negative impact that is largely above that of any single biased human being. The fact that a single AI can negatively and significantly impact the lives and livelihood of millions of people makes increased scrutiny for AIs all the more important. Certain characteristics of AI further exacerbate challenges posed by scale. Many individuals may not be aware that they are being subjected to an AI process, which makes exercising their right to obtain remedies for injuries caused by such an AI very difficult. Additionally there is the "black-box" nature of AIs, and their inherent opacity, with regards to how they work and the data they use, as well as their lack of explainability, which makes the enforcement of current legislation difficult.

The second is purpose, meaning that the existing legislation was not designed to prevent discrimination caused by AI. This gives rise to a two pronged issue. The first prong is that the current legislation's purpose is not clearly focused on preventing discrimination by AI or ensuring accountability of AI developers. For instance, a large part of the European legal framework that aims to prevent discrimination is the GDPR, and yet the GDPR's core purpose is not to protect individuals against discrimination. The second prong is that currently, the question of discrimination, under the GDPR, is supervised by a regulator specialized in data privacy, not in discrimination. For instance, the regulator in charge of enforcing the provisions of the GDPR covering discrimination in France is the "Commission nationale de l'informatique et des libertés" ("CNIL"), which is focused on data privacy issues and must supervise 80,000 entities with a budget of only 18 million euros and 210 staff members, CNIL[119]. Asking whether such a regulator should also be in charge of verifying that millions of job applications are treated in a non-discriminatory manner seems legitimate.

The third and last challenge is the profoundly evolutive nature of this technology, and our inability to tell where it will be at in a few years, let alone decades. This entails that any potential framework regulating AI in recruitment must strike a balance between being sufficiently flexible, as not stifle innovation and remain relevant for years to come, while being sufficiently strict, as to not leave damaging and discriminatory behaviors unpunished.

The In-scope Jurisdictions have legislative frameworks in place to deal with protected characteristics including sex, which have covered traditional gender discrimination in the workplace. Whilst recognizing supranational laws and efforts such as the GDPR and other EU laws as well as the sectoral approach in the US, the patchwork laws and mechanisms for governance, oversight and accountability have failed to keep pace with the use of AI within the employment industry and the related ever evolving and emerging issue.

The GDPR and existing EU and national laws within the In-scope Jurisdictions provide a starting point and can be further supplemented by new national legal frameworks focused on AI in employment, to deal with the unique and evolving challenges presented by the use of AI in recruitment and employment. These frameworks need to be flexible and capable of evolving to address the changing needs of the AI employment sector. There should also be coherence amongst the existing legal frameworks.

In addition to labor law and the GDPR, there exists a broad variety of sector specific legislative frameworks that could either be directly applicable to the issue of discrimination in AI recruitment, or could be used as reference points in designing the future framework regulating AI in recruitment and employment[120].

**119** Direction de l'action du Gouvernement, publications officielles et information administrative, Sénat 146 et 147 (2018-2019).

**120** White Paper On Artificial Intelligence—A European Approach To Excellence And Trust, Eur. Commission (Feb. 19, 2020), https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

Current regulations pertaining to human rights and consumer rights, such as the Directive on equal treatment between men and women in relation to employment[121] and access to goods and services[122], the Consumer Rights Directive[123] or even the Dodd-Frank Wall street Reform and Consumer Protection Act[124] could be used as a legal basis to obtain increased transparency from AIs[125]. The framework used by these regulations could be used to define liability between developers of AI and the company using them, in order to better protect data subjects from discriminatory AI. For example, the Product Liability Directive, passed in 1985, was instrumental in establishing the rules to be followed in allocating responsibility between consumers and product developers in the EU, by introducing the concept of strict liability, whereby producers are responsible for defective products, regardless of whether they are responsible for said defect[126].

The Product Liability Directive was introduced concurrently with other complementary directives, pertaining in particular to product safety[127]. These directives, implemented common safety rules between EU member States, thereby ensuring that companies competing on the EU market all compete with the same rules, and have the same constraints, thus levelling the playing field. It is clear how a similar concept of strict liability, or perhaps shared liability, could be introduced to allow individuals who are victim of a discrimination to easily determine clear liability, and how having shared "safety rules" pertaining to AI and in particular to discrimination can protect data subjects from discriminatory AI, while ensuring a level playing field between developers. Many of the suggestions below draw from these and other existing legal frameworks to recommend a comprehensive legal solution to build upon existing data protection laws and assist with eliminating bias in AI.

**121** Council Directive 2006/54/EC, 2006 O.J. (L 204).

**122** Council Directive 2004/113/EC, 2004 O.J. (L 373).

**123** Council Directive 2011/83/EU, 2011 O.J. (L 304).

**124** Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, 124 Stat. 1376, 1871 (2010), hereinafter Dodd-Frank Act.

**125** Eur. Commission, supra note 120

**126** The Product Liability Directive 85/374/EEC.

**127** The Liability for Defective Products Directive (1999/34/EC).

If designed well, new frameworks for governing the use of automated employment / recruitment systems can provide continued innovation while also articulating clear and predictable guidance for governments, the private sector and civil society organizations to work together for lasting institutional and cultural systems change.

## < 4.1 > Use Disaggregated Data

Data is a critical component of AI. Unfortunately, much of this data is not disaggregated and fails to account for the differences between men and women[128]. There are well documented examples where the failure to account for women can have fatal consequences in medicine, but this failure can also have serious consequences for women in employment decisions that have a ripple effect throughout their entire career[129]. In many cases, aggregated data can perpetuate discrimination particularly if the data labeling is also biased. To avoid these consequences, AI systems should use disaggregated data. Further, gender data collection methods should take into account stereotypes, social and cultural factors that may induce gender bias, gender-proxy language and proper labeling so that we can effectively interpret this data.

## < 4.2 > Extend GDPR coverage to partially automated decisions

As noted above, the GDPR and other legislation does not cover partially automated recruitment processes[130]. Decision-making that is solely automated in the recruitment context is rare, humans are, and will continue

**128** See Caroline Criado Perez, Invisible Women: Data Bias in a World Designed for Men (Abrams Press 2019) (citing several examples of the failure to use sex-disaggregated data in medical, safety and employment decisions).

**129** Id. At 105-111 (arguing the myth of meritocracy may be responsible for the deliberate decision not to use disaggregated data).

**130** See, e.g., The EU Guidelines, supra note 92, at 22 (confirming that the scope of Article 22 should be interpreted extensively: decisions based "solely on automated means" must include any decision in which the human intervention is not meaningful); Gianclaudio Malgieri, Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations, 35(5) Computer L. & Security Rev. (2019), https://www.sciencedirect.com/science/article/pii/S0267364918303753#bcit_16.

to be, involved in at least some of the stages of almost all recruitment processes. While it could be argued that the recruitment process can be split into standalone phases, some phases that are fully-automated and some that are not, it is not clear whether the GDPR would take the different phases of a process into account. Even if it did, the regime would then apply only to certain phases of recruitment through AI.

Removing the current distinction between solely and non-solely automated decision-making in the GDPR would provide simplicity and act as a blueprint for enhanced transparency that could lead to improving the use of AI in employment selection processes and ultimately reduce discrimination. For instance, privacy notices would need to be issued in all cases to make individuals aware of any use of AI to the same degree as required for decision-making that is solely automated i.e. "meaningful information about the logic involved, as well as the significance and envisaged consequences" of such processing[131].

## < 4.3 > Transparency and Accountability

The legal framework targeting AI in recruitment and employment must promote transparency and accountability. It should also include specific provisions to address the risk of bias and discrimination arising from the use of AI in recruitment and employment.

At present, for In-scope Jurisdictions, if an individual sought to bring a discrimination claim stemming from the use of AI, that individual would need certain information to know or suspect that they may have a claim in the first place, such as whether AI was used and how the AI was implemented in the process[132]. However, these AI systems currently operate outside the scope of meaningful scrutiny and accountability. To most employees or

**131** The EU Guidelines, supra note 92, at 20.

**132** The right to request data can be a critical first step in determining whether one has a claim. The UK's Fawcett Society is one of the UK's leading charities campaigning for equal rights of women at work. In November 2019, the Fawcett Society published a report addressing this issue – and recommended that women at work who suspect pay discrimination should have the right to request the salary data of their male colleagues. Andrew Bazely & Gemma Rosenblatt, Why Women Need A Right To Know: Shining A Light On Pay Discrimination 7, Fawcett Soc'y (2019), https://www.fawcettsociety.org.uk/why-women-need-a-right-to-know-shining-a-light-on-pay-discrimination (follow hyperlink).

job candidates, it is unclear how the systems work, what information is entered into them and ultimately how decisions are made. This makes it difficult for job applicants or employees to know whether they may have been discriminated against in the recruitment process or an employment relationship due to their employer's, or prospective employer's, use of AI.

The easiest way to improve transparency around the use of AI, would be to provide access to this information. Indeed, there should be responsible disclosure to ensure people know when they are engaging with AI and what impact such an engagement could have on their employment. For example, the legal framework could follow the example of the State of Illinois in the US and create a law similar to the Artificial Intelligence Video Interview Act, which requires employers to notify applicants in writing that AI may be used to analyze their video interview, explain how the AI works, the characteristics and weightings it uses to evaluate applicants, and obtain the individual's consent to be evaluated by AI[133].

Greater transparency can drive improved compliance with gender discrimination laws by making it easier for individuals to enforce their existing rights. Employers using AI for recruitment should actively disclose and explain to individuals in an understandable manner that they are subject to or contributing data to an AI driven process.

The Council of the European Commission for Human Rights recommends that AI systems be made public in clear and accessible terms, so that individuals are able to understand how decisions are reached and how those decisions have been verified, which aligns with Council of Europe Commissioner for Human Rights Recommendation Unboxing artificial intelligence: 10 steps to protect human rights[134]. The information provided should be focused on the key information required for individuals to enforce those rights. This will include how underlying code is used by the particular program and how a particular decision has been made—technical and non-technical explanations should occur. Such information (particularly in relation to the underlying

**133** Artificial Intelligence Video Interview Act, Ill. Gen. Assembly HB2557 (Aug. 9, 2019), http://www.ilga.gov/legislation/101/HB/PDF/10100HB2557lv.pdf.

**134** Comm'r for Human Rights, Council of Eur., Unboxing Artificial Intelligence: 10 steps to protect Human Rights (2019), https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64=.

code) is likely to be commercially sensitive, so any legal requirement that it be published would need to balance the need for transparency with commercial sensitivity[135]. Current legislation does not provide for this, as demonstrated by the GDPR which, while it establishes several rights for individuals that relate to processing of data, it does not create an express right to an explanation of an AI informed decision.

The right to know that an individual was subject to AI in an employment decision, and to what extent the AI was involved in the decision making process could highlight discriminatory behaviors in AI, and assist individuals in exercising their right to be free from discrimination. While transparency may assist in highlighting potentially discriminating practices and does not in itself directly prevent discriminations, it could be an effective deterrent.

To date, the GDPR has had the most impact of any legislation to create a more regulated data market—and data is the key ingredient for AI systems—however it is now clear that a number of gaps exist within the GDPR with regards to mitigating and correcting for discrimination and bias in AI recruitment and employment. We must ensure that machine learning does not embed and make explicit an already implicitly biased system into our futures. In addition, the GDPR creates no obligation to implement any kind of AI impact assessment or formal disclosures when developing such tools.

### < 4.3.1 > The AI impact assessment

Greater transparency could also be achieved through performing an AI impact assessment, which would apply to both a decision-making process that is either solely automated or non-solely automated, to allow coverage of automated hiring systems. Impact assessments are well-established practices that have been successfully used in scientific and policy domains

**135** Id. (recommending to national authorities, which the Council also states could apply to the private sector, that public authorities should not acquire AI systems from third parties in circumstances where the third party is unwilling to waive restrictions on information—e.g. confidentiality or trade systems—and such systems impede or frustrate the process of carrying out HRIA and marking HRIAs available to the public).

such as environmental protection[136], human rights[137], data protection[138], and privacy[139]. Such impact assessments could be repeated at regular intervals for AI systems that change over time to ensure continued compliance. Testing should account for the origins and use of training data, test data, models, Application Program Interface (APIs), and other components over a product lifecycle. Testing should cover pre-release trials, independent auditing, certification, and ongoing monitoring to test for bias and other harms. While impact assessments are only as good as their uptake and implementation of their findings by decision-makers, Impact assessments have been recognized over the years as forward looking instruments to proactively advise and guide decision-makers on what might happen as well the risks and ways to mitigate risks, such as discrimination and bias.

Impact assessments consider the implications for people and their environment of proposed actions before the action has been undertaken and while there is still the opportunity to modify a proposal (or even abandon the proposal, if appropriate)[140]. An impact assessment provides information for decision-making that analyses social, economic, biophysical and institutional consequences of proposed actions, promotes transparency and participation of the public in decision-making; identifies procedures and

**136** D. Komínková, Environmental Impact Assessment and Application—Part 1, in Encyclopedia of Ecology 1321-1329 (Sven E. Jørgensen & Brian D. Fath eds., 2008) (noting that the Environmental Impact Assessment (EIA) was developed as a tool to minimize the negative impact of human activities on the environment and explaining that the purpose of the EIA is to assess the impact of a proposed activity on the environment before making the decision on whether to carry the activity out); Konrad Ott, Franziska Mohaupt & Rafael Ziegler, Environmental Impact Assessment, in Encyclopedia of Applied Ethics 114-123 (2d Ed. 2012); Environmental Impact Assessment: Theory and Practice 1-2 (Peter Wathern, ed., 1988); Patricia Rodriguez Fortun, Environmental Impact Assessment, European Union Capacity4dev Environment, Climate Change & Green Economy Wiki (last updated Sept. 6, 2017), https://europa.eu/capacity4dev/public-environment-climate/wiki/environmental-impact-assessment.

**137** See, e.g., United Nations, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, U.N. Doc. No. HR/PUB/11/04 (2011), https://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf.

**138** See, e.g., Data Protection Impact Assessments, Information Comm'r's Office (U.K.), supra note 97.

**139** Kenneth A. Bamberger & Deirdre Mulligan, Privacy decision making in Administrative Agencies, U. Chi. L. Rev. 75(1):75 (2008), http://lawreview.uchicago.edu/sites/lawreview.uchicago.edu/files/75_1_Bamberger_Mulligan.pdf.

**140** Júlio de Jesus & Richard Morgan, Impact Assessment, Int'l Ass'n for Impact Assessment (Loreley Fortuny ed., 2020), https://www.iaia.org/wiki-details.php?ID=4 (last visited July 5, 2020).

methods for follow-up (monitoring and mitigation of adverse consequences); and contributes to sustainable development[141]. Impact assessments have been applied to all levels of decision-making from policy design to the individual project level. Various types of impact assessments can be found at Appendix A.

The purpose of the AI impact assessment should be to i) provide transparency with regards to the data used, ii) ensure explainability, by defining the purpose and studying the impact of the AI and iii) provide a scale against which users and developers can be held accountable. In essence, the AI impact assessment aims to prevent the use of discriminatory AI, since if the AI impact assessment concludes that the concerned AI is discriminatory, no corporation will want to sell it, and no corporation will want to use it. Just as the GDPR requires "privacy by design"[142], AI impact assessments should aim for "fairness by design". Studies have shown how a commercial predicting recidivism risk assessment software, COMPAS, widely used in the United States, was biased against African-Americans[143], and how even though it claimed to be using over 137 different features to predict its outcomes, similar results could be found by using only 7 features.

An AI impact assessment may have discovered this before it was deployed and prevented the discrimination from happening in the first instance. The question of fairness by design is central to the fight by women against discrimination in recruitment, but more also to the fight against all groups that might be subject to a discriminatory AI process where the impact can have lifelong negative effects.

---

**141** Id.

**142** Ahmed Abbasi, Jingjing Li, Gari Clifford & Herman Taylor, Make "Fairness by Design" Part of Machine Learning, Harv. Bus. Rev. (Aug. 1, 2018), https://hbr.org/2018/08/make-fairness-by-design-part-of-machine-learning.

**143** Julia Dressel & Hany Farid, The accuracy, fairness, and limits of predicting recidivism, Sci. Advances, Jan. 17, 2018, at 2 (corrected Mar. 30, 2018) (highlighting studies showing that the COMPAS software that widely used in the United States was biased against African-Americans and even though it claimed to be using over 137 different features to predict its outcomes, similar results could be found by using only 7 features). See, also, Julia Angwin et al., Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Moreover, the GDPR's current Data Protection Impact Analysis ("DPIA") regime could be complemented with an Algorithmic Impact Analysis ("AIA") framework, similar to the Canadian Algorithmic Impact Analysis[144]. Unlike the DPIA which only provides broad headings in data protection guidance, an AIA's targeted approach would ensure that companies focus on relevant factors and, once completed the AIA allocates an "impact level" for the proposed use of AI to which proportionate requirements are then attached.

For example, if the intended use of AI has a low level of impact, no peer review of the AI would be required, but if a higher level impact is reached, then more stringent measures could be required, such as peer review by at least two external parties[145]. By specifying the actions to be taken (i.e. different actions depending on impact levels), the AIA could provide unambiguous and tailored protection measures in contrast to the GDPR' s DPIA regime, which requires companies to decide for themselves. In this context, the questions within the AIA must be flexible and capable of evolving to respond to the dynamic needs of technology.

The AIA could also incorporate a Human Rights Impact Analysis ("HRIA"), aligning with the UN Guiding Principle on Business and Human Rights, and with the Council of Europe Commissioner for Human Rights Unboxing artificial intelligence: 10 Steps To Protect Human Rights (Recommendation), which provides a number of steps grounded in international human rights law that national authorities (and the private sector) can take to prevent or mitigate the negative impact AI systems may have on people's lives and rights while also maximizing their potential. It includes recommendations on HRIA in the context of AI, which can be applied by the private and public sectors on AI systems acquired, developed and/or deployed. See Appendix B[146].

**144** Algorithmic Impact Assessment, Gov't of Canada (June 3, 2020), https://canada-ca.github.io/aia-eia-js/ (last visited July 5, 2020).

**145** See, e.g., id. at App'x C.

**146** Comm'r for Human Rights, supra note 134 (listing a number of recommendations rooted in the existing universal, binding and actionable framework provided by the international human rights system that focus on: public consultations; human rights standards in the private sector; information and transparency; independent oversight; non-discrimination and equality; data protection and privacy; freedom of expression; freedom of assembly and association, and the right to work; access to remedies; and the promotion of artificial intelligence).

### < 4.3.2 > Transparency with regards to data and AI system used

Setting aside issues around source code, providing information on the input data, how it was manipulated, labeled and cleaned for the algorithm that was used in the recruitment process, as well as how the type of algorithm and how it was used in the individual decision, could meaningfully contribute to ensuring better predictions and ultimately mitigate discrimination against women[147]. This is underscored by the need to ensure that the issues tied to data bias and label bias be addressed[148].

Understandably, companies developing AI systems will oppose transparency regarding the input data, given that the dtatasets used are valuable assets, accumulated or purchased by such companies and their publication could render them worthless. Therefore, complete transparency does not seem achievable or desirable, but meaningful transparency should be required. Indeed, there is meaningful information that can be provided to the public, for instance there are "fairness" metrics that can be established[149], and even if the dataset itself does not need to be made public, the "fairness" of the dataset used could be made public.

Basic tests, such as representation of protected groups, shouldn't infringe commercial confidentiality and therefore, could be made public.

**147** See Kraft-Buchman & Arian, supra note 15 at 4 (noting that low quality data can be either poorly selected data or incomplete, incorrect or outdated data—could all lead to poor predictions and potentially discrimination against women); European Union Agency for Fundamental Rights, #BigData: Discrimination In Data-Supported Decision-making, FRA (2018), https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf.

**148** See Kraft-Buchman & Arian, supra note Error! Bookmark not defined. (citing several examples where data from studies done on men are applied to women, such as the levels of radiation used for cancer treatments, which have been shown to be unsafe or event deadly for women).

**149** Rachel K.E. Bellamy et al., AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias, Cornell Univ. (Oct. 3, 2018), https://arxiv.org/abs/1810.01943.

### < 4.3.3 > Purpose and Impact

In addition to lacking transparency, companies developing recruitment AIs have no obligation to explain how their tools work, and what were the guiding principles they followed during design development. Moreover, companies using AI in recruitment have no obligation to explain why they are using AI or to determine its impact.

Requiring companies to reveal specific technical information on their design methodology would be controversial, since such information used in designing AI systems could potentially be used by competitors to develop similar products entirely based on the work product of another company. However, we must balance the need to retain corporate confidential information with avoiding harm to the public. Instead, companies could be required to disclose the purpose of their AI system and the approach they had in designing it. This disclosure could be included in the AI impact assessment, in which companies provide answers to simple, yet essential, questions such as "what is the purpose of this AI?" and "what steps have been taken and will be taken to limit bias?".

The scale and repeatability of the potential types of discriminations brought by the use of AI in recruitment, make tools such as an AI impact assessment a cost effective way of avoiding potentially massive discriminatory practices, and will also enforce best practices by forcing companies to answer questions on fairness and discrimination that they might not have thought to address otherwise.

### < 4.3.4 > Disparate Impact Assessments

Disparate impact assessments are used to show how seemingly neutral practices can have hidden negative effects on certain groups[150]. Although it would seem to be in a company's best interest to conduct disparate impact assessments prior to use of these systems, developers and businesses do not appear to undertake such assessments.

**150** Travis Korte & Daniel Castro, Disparate Impact Analysis is Key to Ensuring Fairness in the Age of the Algorithm, Center for Data Innovation (Jan. 20, 2015) available at https://www.datainnovation.org/2015/01/disparate-impact-analysis-is-key-to-ensuring-fairness-in-the-age-of-the-algorithm/

Disparate impact assessments could be embedded into law, and at the very least into the AI impact assessment to provide greater transparency. The existing anti-discrimination statutes, including the UK Equality Act 2010, the French Labor Code and the American Civil Rights Act could be amended to require disparate impact assessments for AI used in employment decisions or alternatively, new laws could be enacted to require such assessments. The law could require that these assessments be provided to the appropriate regulatory body charged with enforcing anti-discrimination laws.

### < 4.3.5 > Accountability

There are a number of challenges to achieving accountability in the use of AI in employment decisions, beginning with the current opaqueness of AI decision-making through the potential removal of humans in fully automated decision-making processes[151], to the use of AI recruitment tools imported from other jurisdictions.

However, accountability is central to addressing these, and other, challenges. Robust accountability tools and frameworks must be developed to assist those who are designing, and deploying AI for work, as well as for those who are the subjects of these AI systems so that we mitigate for discrimination and bias, and provide access to remedies to ensure no one is left behind[152].

Importantly, the legislation should allocate accountability for discrimination resulting from the use of AI in employment decisions[153]. Furthermore, clarifying ambiguity regarding liability in the context of AI-informed recruitment decision-making could contribute to improving accountability. For example, the legal framework targeting employment-related AI could create a rebuttable presumption that the legal person who deploys an employment-related AI system would be liable for any damages caused by

**151** See Australian Human Rights Commission, supra note 44, at 7.

**152** Sci. Foresight Unit (STOA), European Parliamentary Research Serv., A Governance Framework for Algorithmic Accountability and Transparency, Eur. Parliament 55 (Apr. 2019), http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf

**153** See Comm'r for Human Rights, supra note 134, at 10-11; Australian Hum. Rts. Commission, supra note 44, at 8.

using that system[154]. In addition, the question of AI fairness should represent a key part of any new legal framework, allowing candidates, employers and developers to have measurable metrics on which they may assert or refute the discriminatory nature of the AI system.

It is for this purpose that accountability should be differentiated depending on whether the designer or the user is responsible for damages caused by using the AI system. For developers, the accountability mechanism could start with an AI impact assessment. If the AI impact assessment shows that the developers did not sufficiently take into consideration the issue of gender discrimination in the development and implementation of their AI, they could be held accountable for any damages caused by the use of their AI in the employment process.

Similarly, for users of AI, the user could be under a duty to review the AI impact assessment and determine whether the developer gave sufficient consideration to whether there could be a disparate impact on women or other protected classes. Failure to carry out this duty, could result in the company being held responsible for any damages that the use of the AI system caused.

This solution would hold both parties, the developers and the users, accountable when the design of the AI and its use resulted in discriminatory practice.

## < 4.4 > Maintain the Human Element

One the major flaws of certain forms of AI is its lack of transparency and thus, a lack of explainability. Indeed, since it is very difficult to explain the decision-making process of certain forms of AI, it seems key to ensure that humans remain in the decision-making loop where these forms of AI are used, and that those using the AI system have a good understanding of its shortcomings. The role humans can play in overseeing, monitoring and intervening in AI informed and automated decision-making is critically important.

**154** See id. (suggesting there could be a clear signpost of who is responsible for ensuring compliance within the AIA incorporating a HRIA). See also Australian Human Rights Commission, supra note 44, at 8.

### < 4.4.1 > Human-in-the-loop

The level of human involvement within automation may contribute to accountability within the AI system and ultimately in mitigating discrimination and bias[155]. There is evidence that diverse reviewing groups has led to better decision making[156].

New employment legislation could encourage the intervention by diverse groups of human decision-makers in automated decision-making. The legislation could set parameters for this human intervention for example, if the automated decision is at risk of breaching a protected characteristic or would lead to a disparate impact.

### < 4.4.2 > Mandatory Training

Most organizations and individuals that use or develop AI systems are not experts in fairness, explainability and privacy, nor should they be expected to be. Mandatory training, similar to the requirements under the GDPR for an employer to train employees on data protection, could be required and incorporated into the organizational security measures to protect data processing[157].

Such training could target any employee who develops recruitment and employment AI, who inputs data into an AI system or who makes employment decisions based on information provided by AI. The training would inform employees on how the use of AI processes could unintentionally lead to discrimination and enable employees to recognize when information might

**155** Nicol Turner Lee, Paul Resnick, and Genie Barton, Algoritmic bias detection and mitigation; Best practices and policies to reduce consumer harms, Brookings, May 22, 2019, https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms. See Australian Hum. Rts. Commission, supra note 44 at 89.

**156** Jake Silberg and James Manyika, Tackling bias in artificial intelligence (and in humans), McKinsey Global Institute, June 6, 2019, https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans.

**157** See, e.g., General Data Protection Regulation 2018, art. 39. See, also, Council of Europe, Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, 2 (Jan. 2019); Australian Human Rights Commission, supra note 44 at 121.

be biased and act on it. At a minimum, understanding the inherent limitations of AI should be a requirement of being an end-user.[158]

## < 4.5 > Oversight

While creating a system that provides employees the ability to enforce their rights does sometimes drive employers to change their practices, one of the strongest ways to ensure compliance is through the use of an independent body to oversee the actions of the employer. The legislative framework regulating AI in employment could consist of a combination of administrative, judicial, quasi-judicial and/or legislative oversight bodies effectively cooperating with each other. Consideration should be given to empowering, where appropriate, existing National Human Rights Structures (NHRSs)[159] so they can perform a role in providing independent and effective oversight over the human rights compliance of AI systems in employment[160].

The creation of an independent body would also allow for better international coordination and coherence on employment-related AI that is developed in one country and used another. Existing initiatives should also be leveraged. For example, in February 2020 the Organization for Economic Co-operation and Development ("OECD") launched a Policy Observatory on Artificial Intelligence, an online platform to share AI policies[161]. The Policy Observatory on AI aims to facilitate cooperation across OECD countries on policy coherence and help governments develop, implement and improve AI policies by facilitating dialogue and providing multidisciplinary, evidence-based policy analysis on AI[162]. Respective bodies across jurisdictions could coordinate to harmonize their laws.

**158** See Eur. Commission, supra note 92.

**159** Id. at 7.

**160** Id.

**161** See Australian Gov't Dep't of Industry, Science, Energy and Resources, The Global Partnership on Artificial Intelligence launches (June 16, 2020), https://www.industry.gov.au/news-media/the-global-partnership-on-artificial-intelligence-launches; OECD AI Policy Observatory, OECD.AI, https://www.oecd.ai/ (last visited July 6, 2020).

**162** OECD AI Policy Observatory: A Platform to Share and Shape AI Policies, Official Launch Agenda at 1 (Feb. 27, 2020), https://www.oecd.org/going-digital/ai/oecd-ai-observatory-launch-agenda.pdf

### < 4.5.1 > Independent Regulator

An independent regulator could provide effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities[163].

This centralized body could have authority assigned to conduct audits on employment-related AI and charged with new powers to obtain data from developers and companies to discharge this responsibility. In the context of auditing, safety-critical regulated industries such as aerospace and medicine have reliable auditing processes and design controls that have dramatically improved safety (though not without faults). These same procedures could be applied to the use of AI in employment and recruitment. For example, the aerospace industry uses: design checklists; traceability (concerned with the relationships between product requirements, their sources and system design); and Failure Modes and Effects Analysis (FMEA)—methodical and systematic risk management approach that examines proposed design or technology for foreseeable failures[164].

This approach would complete the independent regulation that is already effectively used in the European Economic Area ("EEA") in relation to the GDPR regime, with an overarching regulator at a European level, and local data protection regulatory authorities in each member state of the EEA. In France the CNIL and in the UK, the ICO, are independent regulatory bodies that are responsible for upholding data protection rights. The ICO registers those who process personal information, receives complaints and then issues fines where appropriate (with fines reaching up to €20 million or 4% of total global turnover of the preceding financial year, whichever is higher). The ICO also maintains a public register of enforcement action that it has taken[165]. An independent regulator focuses data controllers on the threat of serious enforcement action and negative publicity, independent from the need for an individual to bring a claim to enforce their rights. This is further

**163** Comm'r for Human Rights, supra note 134, at 10-11.

**164** Inioluwa Deborah Raji, et al., Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing, FAT* '20 (Jan. 27-30, 2020), https://arxiv.org/pdf/2001.00973.pdf

**165** Some Basic Concepts, Information Comm'r's Office (U.K.), https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-data-protection/some-basic-concepts/ (last visited July 5, 2020).

bolstered by the ability for independent regulators to carry out audits on GDPR compliance. The ICO does not, therefore, only react to claims. As a result, organizations across the EEA have been focused on working on their GDPR compliance programs.

There is also framework in the United States that could be adapted to address AI and equal treatment. The U.S. has several regulatory bodies that are tasked with consumer protection. One of the more recently created oversight bodies was a consequence of the 2008 financial crisis. In an effort to bring significant reform to the financial sector, the Dodd–Frank Wall Street Reform and Consumer Protection Act was passed on July 21, 2010[166]. Although legislation aiming to regulate the financial sector might seem unrelated to discrimination in AI employment tools, there are a number of core concepts of the Dodd-Frank Act that could be used to strengthen the legal framework to protect individuals from unintended discrimination.

The Dodd-Frank Act created new independent regulatory agencies such as the Financial Stability Oversight Council ("FSOC") that worked hand in hand with other existing agencies, such as the Federal Reserve, to monitor financial institutions. The idea that successful legislation considered creating new independent agencies, working in tandem with existing agencies, as a useful means to achieving meaningful accountability should be explored when considering how to protect individuals against discrimination through the use of AI.

One of the core pieces of the Dodd-Frank Act was the obligation for certain financial instructions, the systemically important financial institutions ("SIFIs") to prepare "living wills", which are documents that prepare for reorganization and liquidation plans[167] in case the SIFIs become seriously distressed. These living wills include both public and confidential sections and are submitted for review by several regulatory agencies, such as the Federal Reserve Board and the FSOC, and require that that each SIFI take measures to monitor its own business and financial risk.

The ideas behind "living wills" is profoundly tied to the issues mentioned

**166** Dodd-Frank Act supra note 124.

**167** Resolution Plans, 12 C.F.R. § 381.1 (2011). See, also Dodd-Frank Act § 165(d)(1), 12 U.S.C. § 5365(d)(1) ("rapid and orderly resolution" means "reorganization or liquidation of the covered company. . . under the Bankruptcy Code.")

in this paper since it hinges entirely on transparency and explainability. The purpose of the living wills is for a SIFIs to be transparent about their organization and risk to its regulators, to provide explanations as to their functioning and their risk management and to assist the regulators in identifying where additional safeguards could be put in place across the industry.

Similar to AI tools, the contents of these living wills are highly confidential (although brief public summaries are published), and arguments were made against the implementation of living wills based on the fact that it would generate a high degree of risk that key information pertaining to SIFIs would be leaked to the public[168]. However, despite these concerns, the confidential sections of these living wills have remained confidential and are not readily available to the public.

### < 4.5.2 > Penalties for Non-Compliance

Serious penalties for non-compliance could encourage companies and producers of AI to be proactive in avoiding unintended discrimination. In the UK, a recent Women and Equalities Committee Report on enforcement of the Equality Act 2010 explored differing regimes governing sexual harassment in the workplace and the GDPR[169]. The Committee found that organizations spent a great deal of time ensuring compliance with regulations where there are "really stringent regimes that have criminal and civil sanctions"[170], but in the absence of meaningful sanctions, compliance was poor.

Legislative changes and transparency can build public dialogue, educating people on the potential issues and making people aware that the use of AI can have a gender discriminatory effect. It can also encourage better practice. However, this alone is not proactive enough to force employers to deal with gender discrimination in their use of AI. While recognizing that

---

**168** Nizan Geslevich Packin, The Case Against the Dodd-Frank Act's Living Wills: Contingency Planning Following the Financial Crisis, 9 Berkeley Bus. L. J. 1, 29 (2013), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1961626, at 55.

**169** House of Commons and Women & Equalities Comm., Enforcing the Equality Act: the law and the role of the Equality and Human Rights Commission–Tenth Report of Session 2017–19, House of Commons (2019), https://publications.parliament.uk/pa/cm201719/cmselect/cmwomeq/1470/1470.pdf



**170** Id. at 40.

enforcement in the context of AI will be imperfect (given the scale and dynamic nature of technology and the way AI operates), in any regulatory regime, enforcement and the threat of enforcement action are often the strongest driver of change.

## < 4.6 > Professional Bodies and Multidisciplinary Experts

Additionally, the use of independent professional bodies and groups of multidisciplinary experts (rather than mandated regulatory bodies) could also yield positive benefits. For example, the Recruitment and Employment Confederation, one of the main industry bodies for the recruitment industry publishes guidance on best practice in the recruitment industry and sets the standards expected of the recruitment industry in the UK. Although non-binding, guidance from professional bodies can often rapidly go from industry best practice to the "accepted" practice[171]. Additionally, annual reports by oversight bodies offer another effective way to evaluate the effectiveness of measures taken by employers and developers to eliminate bias in AI.

An independent multidisciplinary group of experts could support the development of national AI employment frameworks. The group of experts could consist of stakeholders from, academia, civil society, industry and the community and could provide leadership and advice to the Government on the creation of this new framework[172]. Creating a legal framework supported by a multi-disciplinary group of experts would increase public trust in the new framework and in the AI being used.

A group of experts would enable better coordination between the Government, industry, women and individuals traditionally left behind to further explore the intricacies of what a legal framework would look like, how it would work in practice and ensure efforts were complementary with existing frameworks and initiatives. For example, an expert body could

**171** John Gerard Ruggie, Taking Embedded Liberalism Global: The Corporate Connection, IILJ Working Paper 2003/2 (History and Theory of International Law Series) (2003)(www.iilg.org) (discussing the effects of best practices on an industry).

**172** See, e.g., Australian Human Rights Commission, supra note 44, at 11.

facilitate public consultation in the design of the new laws as well as take responsibility for developing training.

## < 4.7 > Certification

A certification regime could provide a useful additional layer of protection—the GDPR encourages the establishment of data protection certification mechanisms and data protection seals and marks to demonstrate compliance with the GDPR[173]. Those companies that prove they use AI in a responsible way that limits discrimination could be awarded a certificate by an independent third party or a supervisory authority (e.g. a newly appointed independent regulator or an existing regulatory authority charged with enforcing anti-discrimination laws)[174].

The use of certification regimes is a way to verify that organizations comply with certain rules, such as absence of bias or discrimination. The needs and risk vary greatly between sectors, therefore certification requirements could be sector specific to be most effective[175]. Certification may encourage more careful use of AI by ensuring a company's use of AI forms part of its public image. Certain companies might be willing to obtain such certification in an effort to lower their risk profile, as compliance with a recognized certification regime could be used as a defense to a discrimination claim arising from the use of AI. Such a certification regime could operate similarly to and draw upon the United Nations Development Program's ("UNDP") Gender Equality Seal for Public and Private Organizations, which recognizes public and private

**173** Australian Human Rights Commission, supra note 44, at 112; Eur. Commission, supra note 92; Alan Finkel, Chief scientist Alan Finkel: Why we need a 'Turing Stamp' for AI organizations and products, SmartCompany (May 21, 2018), https://www.smartcompany.com.au/startupsmart/startupsmart-technology/chief-scientist-alan-finkel-we-need-turing-stamp-ai-organisations-products/ (proposing the establishment of a 'Turing Stamp'—a voluntary certification scheme for ethical AI that would be independently audited).

**174** General Data Protection Regulation 2018, art. 42-43 (encouraging the establishment of data protection certification mechanisms and noting that a supervisory authority such as the ICO or CNIL would be a valid certification bodies).

**175** Sci. Foresight Unit (STOA), European Parliamentary Research Serv., Understanding Algorithmic Decision-making: Opportunities and Challenges, Eur. Parliament 4 (Mar. 2019), http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261(ANN1)_EN.pdf.

organizations for meeting specific standards to promote gender equality[176].

There are a number of initiatives aiming to develop technical standards and certification schemes, and existing efforts should be leveraged and utilized, including the promising Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) by the Institute of Electrical and Electronics Engineers ("IEEE") that aims to create specifications for certification and marking processes to improve transparency and accountability, and help to reduce algorithmic bias in AI.[177]

## < 4.8 > International Cooperation

Where national rules may fall short, international standards provide a path towards effective global solutions[178]. Standards can influence the development and deployment of particular AI systems through product specifications such as explainability, robustness, and fail-safe design, and through process specifications can affect the larger context in which AI is researched, developed, and deployed[179]. Standards development is underway

**176** Gender Equality Seal for Public and Private Organizations, UNDP, https://www.undp.org/content/undp/en/home/2030-agenda-for-sustainable-development/people/gender-equality/transforming-workplaces-to-advance-gender-equality/gender-equality-seal-for-public-and-private-enterprises.html (last visited July 5, 2020) (providing partners with tools, guidance and specific assessment criteria to ensure successful implementation and Gender Equality Seal certification evidencing a company supports a more efficient and equitable workplace and contributes to the advancement of gender equality and the achievement of the Sustainable Development Goals).

**177** Press Release, IEEE Standards Ass'n, IEEE Invites Companies, Governments and Other Stakeholders Globally to Expand on Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) Work (Feb. 26, 2020), https://standards.ieee.org/news/2020/phase-1-ecpais.html (noting that ECPAIS was launched at the end of 2018 by IEEE-Standard Association, in collaboration with founding partners including the Finnish Ministry of Finance, the cities of Espoo, Vienna, and New York, UN bodies such as UNICEF, and the UN Secretary General's High-level Panel on Digital Cooperation, and industry partners such as EY and Accenture).

**178** GLOBAL RESEARCH COUNCIL, SUPPORTING WOMEN IN RESEARCH: POLICIES, PROGRAMS AND INITIATIVES UNDERTAKEN BY PUBLIC RESEARCH FUNDING AGENCIES (Apr. 2019), https://anr.fr/fileadmin/documents/2019/GRC_GWG_Case_studies_final.pdf

**179** Peter Cihon, Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development, Future of Human. Inst. 2 (Apr. 2019), https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf.

at ISO/IEC, and IEEE. Over 18 countries are currently developing national AI strategies, and almost all include some provision for the development of standards or approaches, with China and the US prioritizing engagement in the standardization processes for AI[180]. However, it is unclear how many of these potential national standards currently look at discrimination and bias mitigation and correction.

There are several new initiatives relating to the use of AI around the world that are siloed and uncoordinated[181]. Wherever possible there should be coordination internationally to ensure standardization efforts and that the development of any legal frameworks are complementary and grounded in human rights. In this context, the outcomes of the Ad Hoc Committee on Artificial Intelligence ("CAHAI"), established by the Council of Europe should be leveraged[182]. The CAHAI's mandate is to examine the feasibility and potential elements, based on broad multi-stakeholder consultations, of a legal framework for the development, design and application of AI, based on Council of Europe's standards on human rights, democracy and the rule of law. The CAHAI's mandate states that when fulfilling this task, the CAHAI should take due account of a gender perspective[183]. International cooperation and multi-stakeholder cooperation of governments, civil society, academics, technologists and the private sector.

**180** Id. at 21-22, 25.

**181** Sci. Foresight Unit (STOA), supra note 152 (noting E-commerce proposals being discussed at the WTO and regional trade negotiations that include proposals to protect Intellectual Property by restricting access to information regarding proprietary algorithms).

**182** Ad Hoc Committee on Artificial Intelligence – CAHAI, Council of Eur. (2020), https://www.coe.int/en/web/artificial-intelligence/cahai.

**183** Ad hoc Committee on Artificial Intelligence (CAHAI): Terms of Reference. Main Tasks, 1353rd Meeting, Comm. of Ministers, Council of Europe (Sept. 11, 2019), Council of Eur., https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016809737a1 ("The Committee shall take into account the standards of the Council of Europe relevant to the design, development and application of digital technologies, in the fields of human rights, democracy and the rule of law, in particular on the basis of existing legal instruments; take into account relevant existing universal and regional international legal instruments, work undertaken by other Council of Europe bodies as well as ongoing work in other international and regional organisations; and take due account of a gender perspective, building cohesive societies and promoting and protecting rights of persons with disabilities in the performance of its tasks.")

Diverse groups would be beneficial in developing new regimes. There could be increased benefit by including these groups[184].

The mass scale correction of skewed data systems (a known contributor to the root cause of discrimination and bias in AI)[185] will require multilateral and international cooperation to ensure we leave no one behind. The UN is positioned well to lead on this, for example through putting the gender data gap on the official agenda of its third UN World Data Forum in Switzerland in 2020, and coordinating with existing efforts[186]. It is envisaged that this action would be grounded in human rights, and guide and provoke creative thinking on a whole system approach for closing the gender data gap that is fit for that purpose in this fast-changing digital age[187].

In April 2019, the EU Commission presented the Ethics Guidelines for Trustworthy Artificial Intelligence, which is currently going through a piloting process[188]. While not mandatory, these Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy. These requirements could also serve as a basis for a strong supranational legal framework for AI in employment. They include: (i) human agency and oversight; (ii) technical robustness safety; (iii) privacy and data governance; (iv) transparency; (v) diversity, non-discrimination and fairness; (vi) societal and environmental well-being; and (vii) accountability.

**184** The Age of Digital Interdependence: Report of the UN Secretary-General's High-level Panel on Digital Cooperation 4 (2019), https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf

**185** Amina Adad & Mohammed Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (XAI), IEEE Access, 6, 52138–52160 (2018) (available at https://ieeexplore.ieee.org/document/8466590).

**186** See Kraft-Buchman & Arian, supra note 17, at 17-18 (highlighting that the outcomes of UN Women's flagship programs, Making Every Woman and Girl Count and Flagship Program Initiative: Better Production and Use of Gender Statistics for Evidence-Based Localization of the SDGs (Gender Statistics FPI) and recently launched PARIS21 comprehensive framework to assess data and capacity gaps linked to gender statistics should also be leveraged).

**187** Id.

**188** Eur. Commission, supra note 86.

< 5 >

# Conclusion

Frameworks that foster an environment where women and girls can achieve full participation and equal enjoyment of rights are needed. As noted above gender discrimination in an employment context, whether perpetrated by a human or through AI bias, is unlawful in the In-scope Jurisdictions of this report.

Despite this, there is little existing legislation addressing the use of AI in employment decisions—existing employment law may fall short of being able to detect and therefore ultimately prevent discrimination arising from the use of AI. Although there has been a slow reaction to the mounting evidence of gender bias in employment-related AI systems, there are things we can do now to be proactive in preventing discrimination.

COVID-19 will affect the global economy in ways that are unforeseen. One way to strengthen societies and our economy is to inoculate against the bias in AI systems that can exclude or block large amounts of desperately needed human talent. A course correction will enable us to leverage the untapped innovation and creativity of women and those traditionally excluded in the workforce. But to do that we must act now.

< 6 >

# Recommendation shortlist

Governments, the private sector and civil society organizations can create a gender responsive digital landscape to advance the values of equality we have long embraced and to correct for the visibility and influence of women proportionate to the population. In addition to the steps being taken to promote an intersectional variety and equal number of women and girls in decision-making positions and in the creation, design and coding of AI in recruitment and employment, a strong legal framework is needed to address the dynamic needs of the employment and recruitment industry and should:

> **Facilitate gender equality in data by using data that is disaggregated** and that responsively addresses gender insights, and that includes the concept of fairness by design.

> **Strengthen the GDPR** by removing the distinction currently made between solely and non-solely automated decision-making, and strengthening transparency and explainability requirements.

> **Incorporate the Use of Algorithmic Impact Assessments** that include a Human Rights Impact Assessment ("**HRIA**") of employment-related AI systems.

> **Establish and coordinate with independent regulatory oversight bodies** for the use of AI in recruitment and employment.

> **Require human involvement** in all recruitment automation systems designed or used in order to facilitate human intervention in all AI-informed decision-making, giving the human decision-maker the ability to override any automated decision if there is a risk of discrimination or bias.

- **Build capacity through mandatory training** for any individual who undertakes AI-informed decision-making or develops or inputs data into a recruitment automation system.

- **Establish a certification regime** whereby an independent third party could award recruitment certification for organizations that put gender equality at the center of the design / use of AI in recruitment and employment.

- **Encourage multilateral and international cooperation grounded in human rights,** to correct the mass scale skewed data systems and coordination amongst AI standardization efforts to give all of these recommendations more power and coherence.

< Appendix A >

# Examples of Impact assessments

## < 1 > Environmental Impact Assessment

Environmental Impact Assessments ("**EIA**") are now universally recognized instruments for environmental management. 191 of the 193 member nations of the United Nations either having national legislation or having signed some form of international legal instrument that refers to the use of EIAs[189].

The well-developed support infrastructure from professional groupings (such as the International Association of Impact Assessments and its national affiliates and branches) to support from international agencies (such as UN Environment, the World Bank Group and World Health Organization) as well as national environmental agencies and territory institutions (providing guidance material and other resources) and the engagement of a community of researchers and practitioners (through case studies and theory-based analysis) has contributed to the success of EIAs in impacting decision-making at the highest levels, for almost fifty years now. There are consistent and regular efforts to ensure EIAs remain effective within the context it is used.

**189** R ichard K. Morgan, Environmental Impact Assessment: The State of the Art, 30 Impact Assessment & Project Appraisal 5, 6 (2012), https://www.tandfonline.com/doi/pdf/10.1080/14615517.2012.661557?needAccess=true.

## < 2 > Human Rights Impact Assessments (HRIA)

In order for companies to exercise their human rights due diligence they must: assess actual and potential impacts including through HRIA; integrate and act upon findings; track performance; and communicate how the organization is addressing actual and potential impacts[190]. HRIA's are an important tool for mitigating and addressing adverse human rights impacts and could be undertaken ahead of using AI for talent acquisition. HRIA is a process for identifying, understanding, assessing and addressing the adverse effects of programs, projects and activities on the human rights enjoyment of workers, communities, consumers or other rights-holders[191].

According to the UN Guiding Principles on Business and Human Rights, businesses should do the following when assessing their human rights impacts:[192]

–Draw on internal and/or independent human rights expertise;
– Undertake meaningful consultation with potentially affected rights-holders and other relevant parties;
– Be gender-sensitive and pay particular attention to any human rights impacts on individuals from groups that may be at heightened risk of vulnerability or marginalization;
– Assess impacts from the perspective of risk to people rather than risk to business; and
– Repeat their risk and impact identification and assessment at regular intervals (i.e. before entering into a new activity, prior to significant decisions about changes in activities, and periodically throughout the project-cycle).

**190** See United Nations, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, U.N. Doc. No. HR/PUB/11/04 (2011), https://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf (elaborating on the implications of existing standards and practices for states and businesses, to meet their responsibility to protect human rights, and including points covered variously in international and domestic law, including encouraging companies to "know and show" that they do not infringe on human rights through their operations or business relationships and to adopt a human rights policy and to carry out human rights due diligence).

**191** Human Rights Impact Assessments, Nat'l Action Plans on Bus. & Human Rights, https://globalnaps.org/issue/human-rights-impact-assessments/ (last visited July 5, 2020).



**192** Id.; see also Guiding Principles, supra note 190, at 19-20.

Examples of HRIA approaches that have been developed include: "stand-alone" HRIA (i.e. assessments that focus exclusively on human rights); "integrated' assessments" (e.g. integrating human rights into environmental, social and health impact assessments); and others (e.g. community-led, sector-wide, or in the area of trade and investment). A sector-wide impact assessment (SWIA), which goes beyond a particular project to cover a sector as a whole and address project level, cumulative level and sector level impacts holistically is another form of a HRIA[193]. Organizations should carefully select the most appropriate HRIA methodology to cater to the size and complexity of its operations to ensure a HRIA does not become a tick box exercise with limited impact on the human rights of the most vulnerable and / or disadvantaged. To facilitate HRIAs consistent with the UNGPs, the Danish Institute for Human Rights has developed a HRIA Guidance and Toolbox[194].

## < 3 > Algorithmic Impact Assessment and Directive

In Canada, the Directive on Automated Decision-Making[195] (the "**Canadian Directive**") has sought to achieve greater transparency in algorithmic accountability. The Canadian Directive applies to federal institutions (i.e. not private companies) that use AI and requires them to: (i) provide a notice on their website before a decision is being made by an automated decision system; (ii) provide a meaningful explanation after a decision to impacted individuals as to how and why a particular decision was made; and (iii) make public custom source code owned by the Government of Canada and used in the automated decision-making[196]. These measures seek to enhance transparency in this area—and greater transparency is likely to increase the ability of individuals impacted by a particular decision to enforce their existing rights not to be the subject of unlawful gender discrimination.

**193** Human Rights Impact Assessments, supra note 191.

**194** See generally Danish Inst. for Human Rights, Human Rights Impact Assessment: Guidance and Toolbox (2016), https://www.humanrights.dk/sites/humanrights.dk/files/media/dokumenter/business/hria_toolbox/hria_guidance_and_toolbox_final_may22016.pdf_223795_1_1.pdf.

**195** Directive on Automated Decision-Making, Gov't of Canada (Feb. 5, 2019), https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

**196** Id.

Additionally in 2019, the Government of Canada became the first country in the world to develop an Algorithmic Impact Assessment (AIA)[197]. The mandatory AIA requires federal institutions to answer several questions about how the programs will be used and what safeguarding measures have been put in place—prior to the production of any system, tool, or statistical model used to recommend or make an administrative decision about a client[198].

Questions in the AIA are centered on: the business driver /positive impact; about the system (e.g. will it replace a human decision); about the decision; impact assessment (e.g. impact on rights or freedoms, economy and environment); about the data (e.g. will personal information be used); consultations (e.g. will consultations take place with external stakeholders, such as civil society, academia and industry); de-risking and mitigation measures relating to data quality (e.g. has a gender based analysis of the data of the data has occurred); and de-risking and mitigating measures relating to procedural fairness (e.g. will the system provide an audit trail that records all the recommendations or decisions made by the system)[199]. The final results of any AIA conducted have to be released in an accessible format on a Government of Canada website[200]. The AIA aims to produce a comprehensive public record of information around algorithmic decision-making.

Once the AIA questionnaire is completed, Government departments receive an impact level which determines which requirements will apply to the automated decision-making system (i.e. peer review, notice, human-in-the-loop for decisions, explanation requirement, testing, monitoring and training). Impact levels are 1-4: no impact; moderate impact; high impact; and very high impact. Requirements scale according to the risk level; i.e. peer review requirement for level 1 initiatives will be much less than those required for other levels[201].

**197** Algorithmic Impact Assessment, Gov't of Canada, supra note 144.

**198** Id.

**199** Id.

**200** Directive on Automated Decision-Making, Gov't of Canada, supra note 195, art. 6.1.4.

**201** Id. apps. B, C.

< Appendix B >

# AI specific recommendations for Impact Assessments

The Council of Europe Commissioner for Human Rights makes a number of specific recommendations on HRIA in the context of AI, which can be applied to employment-related AI[202]:

- Include a meaningful external review of AI systems, either by an independent oversight body or an external researcher/auditor with relevant expertise, in order to help discover, measure and/or map human rights impacts and risks over time (public bodies should consider involving National Human Rights Structures).

- Self-assessments and external reviews should include an evaluation of how decision-makers might collect or influence the inputs and interpret the outputs of such a system as well as whether an AI system remains under meaningful human control throughout the AI system's lifecycle.

- Where the self-assessment or external review discloses that the AI system poses a real risk of violating human rights, the HRIA must set out the measures, safeguards, and mechanisms envisaged for preventing or mitigating that risk (if a AI system has already been deployed, it should be immediately suspended until safeguards and mechanisms have been adopted). Where it is not possible to meaningfully mitigate the identified risk, the AI system should not be deployed or otherwise used.

- Where the self-assessment or external review discloses a violation of human rights, the public authority must act immediately to address and remedy the violation and adopt measures to prevent or mitigate the risk of such a violation occurring again.

202 Comm'r for Human Rights supra note 134.

- The HRIAs, including research findings or conclusions from the external review process, must be made available to the public in an easily accessible and machine-readable format.

- AI systems should not be acquired from third parties in circumstances where the third party is unwilling to waive restrictions on information (e.g. confidentiality or trade secrets) where such restrictions impede or frustrate the process of (i) carrying out HRIAs and (ii) making HRIAs available to the public.

- Conduct HRIAs on a regular basis, and not only at the point of acquisition and/or development of AI systems. HRIAs should, at the very least, be undertaken at each new phase of the AI system lifecycle and at similarly significant milestones.

# < Bibliography >

### Books and Articles

– Ahmed Abbasi, Jingjing Li, Gari Clifford & Herman Taylor, Make "Fairness by Design" Part of Machine Learning, Harv. Bus. Rev. (Aug. 1, 2018), https://hbr.org/2018/08/make-fairness-by-design-part-of-machine-learning

– Alan Finkel, Chief scientist Alan Finkel: Why we need a 'Turing Stamp' for AI organizations and products, SmartCompany (May 21, 2018), https://www.smartcompany.com.au/startupsmart/startupsmart-technology/chief-scientist-alan-finkel-we-need-turing-stamp-ai-organisations-products/

– Amina Adad & Mohammed Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (XAI), IEEE Access, 6, 52138–52160 (2018) (available at https://ieeexplore.ieee.org/document/8466590)

– Andrew Bazely & Gemma Rosenblatt, Why Women Need A Right To Know: Shining A Light On Pay Discrimination 7, Fawcett Soc'y (2019), https://www.fawcettsociety.org.uk/why-women-need-a-right-to-know-shining-a-light-on-pay-discrimination

– Aurélien Rayer, Exploring Collaborative Filtering For Job Recommendations, Welcome to the Jungle (Feb. 1, 2019), https://www.welcometothejungle.com/en/articles/collaborative-filtering-job-recommendations

– Aylin Caliskan, Joanna J. Bryson & Arvind Narayanan, Semantics Derived Automatically from Language Corpora Contain Human-Like Biases, 356 SCIENCE 183 (2017), https://science.sciencemag.org/content/356/6334/183

– Ben Dattner, Tomas Chamorro-Premuzic, Richard Buchband & Lucinda Schettler, The Legal and Ethical Implications of Using AI in Hiring, Harv. Bus. Rev. (Apr. 25, 2019), https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring

– Caitlin Kraft-Buchman & Renée Arian, women@thetable, The Deadly Data Gap: Gender & Data 3-8 (2019), available at https://uploads.strikinglycdn.com/files/cb0104a3-7f50-49c7-858e-827169cceabd/FINAL%20UPLOAD%20The

– Caitlin Kraft-Buchman & Renée Arian, women@thetable, Triple A: Affirmative Action for Algorithms - Artificial Intelligence, Automated Decision-Making & Gender 3-7 (2019), available at https://uploads.strikinglycdn.com/files/6b4e02b8-5405-4087-927c-221a6b0608c8/WeShapeOurTools%20v11.pdf (citing several examples of gender bias in AI)

– Caroline Criado Perez, Invisible Women: Data Bias in a World Designed for Men (Abrams Press 2019)

– D. Komínková, Environmental Impact Assessment and Application—Part 1, in Encyclopedia of Ecology 1321-1329 (Sven E. Jørgensen & Brian D. Fath eds., 2008)
– David Braue, ACS Information Age, Recruiters Tap AI for Post-COVID Staff Hunt, ACS (May 5, 2020), https://ia.acs.org.au/article/2020/recruiters-tap-ai-for-post-covid-staff-hunt.html

– David R. Fineman, People analytics: Recalculating the route: 2017 Global Human Capital Trends, https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2017/people-analytics-in-hr.html

– Devah Pager & Hana Shepherd, The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets, 34 ANN. REV. SOC. 181, 182 (2008)

– Drew Harwell, A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job, Wash. Post, Nov. 6, 2019, https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job

– Elizabeth Levy Paluck of Princeton and Donald P. Green of Yale in the Annual Review of Psychology (2009)

– Final Proposed California Consumer Privacy Act Regulations Submitted for Administrative Review, Privacy & Cybersecurity Update (Skadden, Arps, Slate, Meagher & Flom LLP New York, N.Y.), June 2020

– Forbes Coaches Council, 10 Ways Artificial Intelligence Will Change Recruitment Practices, Forbes (Aug. 10, 2018, 7:00 AM), https://www.forbes.com/sites/forbescoachescouncil/2018/08/10/10-ways-artificial-intelligence-will-change-recruitment-practices/#68b43cc13a2c

– Gianclaudio Malgieri, Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations, 35(5) Computer L. & Security Rev. (2019), https://www.sciencedirect.com/science/article/pii/S0267364918303753#bcit_16

– Golnoosh Farnadi et al., Computational Personality Recognition in Social Media, CORE 109-142, https://core.ac.uk/download/pdf/55893069.pdf

– Grant Thornton, Women in Business: Building a Blueprint for Action 2, 5-6 (Mar. 2019), https://www.grantthornton.global/globalassets/global-insights---do-not-edit/2019/women-in-business/gtil-wib-report_grant-thornton-spreads-low-res.pdf

– Ian Thomsen, Your Gender and Race Might Be Determining Which Facebook Ads You See, News@Northeastern (Apr. 11, 2019), https://news.northeastern.edu/2019/04/11/a-northeastern-university-team-tested-facebooks-algorithm-and-found-its-delivery-of-advertisements-is-skewed-by-race-and-gender/

– Inioluwa Deborah Raji, et al., Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing, FAT* '20 (Jan. 27-30, 2020), https://arxiv.org/pdf/2001.00973.pdf

– Iris Bohnet, Real Fixes for Workplace Bias, Wall St. J. (Mar. 11, 2016), https://www.wsj.com/articles/real-fixes-for-workplace-bias-1457713338
Iris Bohnet, What Works: Gender Equality By Design (Belknap Press of Harv. Univ. Press, 2016)

– Jake Silberg and James Manyika, Tackling bias in artificial intelligence (and in humans), McKinsey Global Institute, June 6, 2019, https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans

– Javier Sánchez-Monedero, Lina Dencik & 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, in ACM, FAT 2020: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, arXiv (Jan. 17, 2020), https://arxiv.org/pdf/1910.06144.pdf

– Jeanne Meister, Top 10 HR Trends That Matter Most in the 2020 Workplace, Forbes (Jan. 15, 2020 11:07 AM), https://www.forbes.com/sites/jeannemeister/2020/01/15/top-10-hr-trends-that-matter-most-in-the-2020-workplace/#320fe6a37dfc

– John Gerard Ruggie, Taking Embedded Liberalism Global: The Corporate Connection, IILJ Working Paper 2003/2 (History and Theory of International Law Series) (2003)(www.iilg.org)

– Josh Bersin & Tomas Chamorro-Premuzic, New Ways To Gauge Talent And Potential, MIT Sloan Management Rev. (Nov. 16, 2018), https://sloanreview.mit.edu/article/new-ways-to-gauge-talent-and-potential

– Julia Angwin et al., Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

– Julia Dressel & Hany Farid, The accuracy, fairness, and limits of predicting recidivism, Sci. Advances, Jan. 17, 2018, at 2 (corrected Mar. 30, 2018)

– Julian Kolev et al., Is Blinded Review Enough? How Gendered Outcomes Arise Even Under Anonymous Evaluation, National Bureau of Economic Research Working Paper Series (2019), https://gap.hks.harvard.edu/blinded-review-enough-how-gendered-outcomes-arise-even-under-anonymous-evaluation

– Júlio de Jesus & Richard Morgan, Impact Assessment, Int'l Ass'n for Impact Assessment (Loreley Fortuny ed., 2020), https://www.iaia.org/wiki-details.php?ID=4 (last visited July 5, 2020)

– Julius Schulte, AI-Assisted Recruitment Is Biased. Here's How To Make It More Fair, World Econ. F. (May 9, 2019), https://www.weforum.org/agenda/2019/05/ai-assisted-recruitment-is-biased-heres-how-to-beat-it

– Kenneth A. Bamberger & Deirdre Mulligan, Privacy decision making in Administrative Agencies, U. Chi. L. Rev. 75(1):75 (2008), http://lawreview.uchicago.edu/sites/lawreview.uchicago.edu/files/75_1_Bamberger_Mulligan.pdf

– Konrad Ott, Franziska Mohaupt & Rafael Ziegler, Environmental Impact Assessment, in Encyclopedia of Applied Ethics 114-123 (2d Ed. 2012); Environmental Impact Assessment: Theory and Practice 1-2 (Peter Wathern, ed., 1988); Patricia Rodriguez Fortun, Environmental Impact Assessment, European Union Capacity4dev Environment, Climate Change & Green Economy Wiki (last updated Sept. 6, 2017), https://europa.eu/capacity4dev/public-environment-climate/wiki/environmental-impact-assessment

– Kumba Sennaar, Machine Learning For Recruiting And Hiring—6 Current Applications, Emerj (May 20, 2019), https://emerj.com/ai-sector-overviews/machine-learning-for-recruiting-and-hiring

– Kyle Wiggers, AI Weekly: Coronavirus Spurs Adoption of AI-powered Candidate Recruitment and Screening Tools, VentureBeat (Mar. 13, 2020), https://venturebeat.com/2020/03/13/ai-weekly-coronavirus-spurs-adoption-of-ai-powered-candidate-recruitment-and-screening-tools/

– Larry Hardesty, Study Finds Gender And Skin-Type Bias In Commercial Artificial-Intelligence Systems, MIT News (Feb. 11, 2018), http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212

– Linda Qu, 99% of Fortune 500 Companies use Applicant Tracking Systems (ATS)
Manish Raghavan & Solon Barocas, Report, Big Data's Disparate Impact 104 Cal. L. Rev. 671, 691 (2016)

– Manish Raghavan & Solon Barocas, Report, Challenges For Mitigating Bias In Algorithmic Hiring, Brookings (Dec. 6, 2019), https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/

– Mary Olson, Atomic radiation is more harmful to women, Wise (Nov. 11. 2011), https://www.wiseinternational.org/nuclear-monitor/736/atomic-radiation-more-harmful-women

– Mary Olson, Females Exposed to Nuclear Radiation Are Far Likelier Than Males to Suffer Harm, PassBlue (July 5, 2017), https://www.passblue.com/2017/07/05/females-exposed-to-nuclear-radiation-are-far-likelier-than-males-to-suffer-harm/

– Melissa Hart, Subjective Decision making and Unconscious Discrimination, 56 Ala. L. Rev. 741 (2005), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=788066

– Michael Rovatsos, Brent Mittelstadt & Ansgar Koene, United Kingdom Center for Data Ethics and Innovation, Landscape Summary: Bias In Algorithmic Decision-Making 19-22 (2019), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf

– Miranda Bogen, All The Ways Hiring Algorithms Can Introduce Bias, Harv. Bus. Rev. (May 6, 2019), https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias

– Nicol Turner Lee, Paul Resnick, and Genie Barton, Algoritmic bias detection and mitigation; Best practices and policies to reduce consumer harms, Brookings, May 22, 2019, https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms

– Nizan Geslevich Packin, The Case Against the Dodd-Frank Act's Living Wills: Contingency Planning Following the Financial Crisis, 9 Berkeley Bus. L. J. 1, 29 (2013), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1961626, at 55

– Nuri Purswani, Word2Vec For Talent Acquisition, Medium.com (Sept. 13, 2018) https://medium.com/@nuripurswani/word2vec-for-talent-acquisition-ab20a23e01d8

– Pauline Kim & Sharion Scott, Discrimination in Online Employment Recruiting, St. Louis Univ. L.J., 63(1) (2018), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3214898

– Peter Cihon, Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development, Future of Human. Inst. 2 (Apr. 2019), https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf.

– Rachel K.E. Bellamy et al., AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias, Cornell Univ. (Oct. 3, 2018), https://arxiv.org/abs/1810.01943 33

– Rebecca Cassells & Alan Duncan, Bankwest Curtin Economics Centre, Gender Equity Insights 2019: Breaking Through the Glass Ceiling, 8, BCEC | WGEA Gender Equity Series 4 (Mar. 2019), available at https://bcec.edu.au/assets/2019/02/BCEC-WGEA-Gender-Equity-Insights-2019-Report.pdf

– Reuben Binns & Valeria Gallo, Automated Decision-making: The Role of Meaningful Human Reviews, Computers & L., June 2019, at 29, 30-31

– Richard K. Morgan, Environmental Impact Assessment: The State of the Art, 30 Impact Assessment & Project Appraisal 5, 6 (2012), https://www.tandfonline.com/doi/pdf/10.1080/14615517.2012.661557?needAccess=true

– Richard Landers, Developing a theory of gamified learning: Linking serious games and gamification of learning, 45(6) Simulation & Gaming 752 (2014), available at https://www.researchgate.net/publication/268632276_Developing_a_Theory_of_Gamified_Learning

– Stuart D. Levi, Skadden's 2019 Insights: California Privacy Law: What Companies Should Do to Prepare in 2019 (Jan. 17, 2019)

– Tomas Chamorro-Premuzic, Dave Winsborough, Ryne A. Sherman & Robert Hogan, New Talent Signals: Shiny New Objects or a Brave New World?, 9 Indus. & Org. Psychology 621, (2016) available at https://pdfs.semanticscholar.org/83ef/b5454fd547d0284c89b2cbf0940ac14459b2.pdf

– Travis Korte & Daniel Castro, Disparate Impact Analysis is Key to Ensuring Fairness in the Age of the Algorithm, Center for Data Innovation (Jan. 20, 2015) available at https://www.datainnovation.org/2015/01/disparate-impact-analysis-is-key-to-ensuring-fairness-in-the-age-of-the-algorithm/

– W. Later et al., Is the 1975 Reference Man Still a Suitable Reference?, Eur. J. Clin. Nutr. 64(10), 1035-42 (2010), https://www.nature.com/articles/ejcn2010125.pdf

## Reports, Guidance, Web Pages and Other Sources

– Ad Hoc Committee on Artificial Intelligence – CAHAI, Council of Eur. (2020), https://www.coe.int/en/web/artificial-intelligence/cahai

– Ad hoc Committee on Artificial Intelligence (CAHAI): Terms of Reference. Main Tasks, 1353rd Meeting, Comm. of Ministers, Council of Europe (Sept. 11, 2019), Council of Eur., https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016809737a1

– Algorithmic Impact Assessment, Gov't of Canada (June 3, 2020), https://canada-ca.github.io/aia-eia-js/ (last visited July 5, 2020)

– Australian Gov't Dep't of Industry, Science, Energy and Resources, The Global Partnership on Artificial Intelligence launches (June 16, 2020), https://www.industry.gov.au/news-media/the-global-partnership-on-artificial-intelligence-launches; OECD AI Policy Observatory, OECD.AI, https://www.oecd.ai/ (last visited July 6, 2020)

– Australian Human Rights Commission, Human Rights And Technology: Discussion Paper December 2019 (2019), available at https://tech.humanrights.gov.au/sites/default/files/2019-12/TechRights2019_DiscussionPaper.pdf

– CJEU Judgment: Police Rule on Height Limit May Discriminate Against Women, EUR-LEX, https://eur-lex.europa.eu/content/news/CJ_police_height.html (last visited July 5, 2020)

– Comm'r for Human Rights, Council of Eur., Unboxing Artificial Intelligence: 10 steps to protect Human Rights (2019), https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64=

– Congressional Research Service, Data Protection Law: An Overview, R45631 (Mar. 25, 2019), available at https://crsreports.congress.gov/product/pdf/R/R45631

– Council of Europe, Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, 2 (Jan. 2019)

– Counsel of Europe, https://www.coe.int/en/web/artificial-intelligence/glossary

– Danish Inst. for Human Rights, Human Rights Impact Assessment: Guidance and Toolbox (2016), https://www.humanrights.dk/sites/humanrights.dk/files/media/dokumenter/business/hria_toolbox/hria_guidance_and_toolbox_final_may22016.pdf_223795_1_1.pdf

– Data Protection Impact Assessments, Information Comm'r's Office (U.K.), https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/ (last visited June 9, 2020)

– Data Protection Working Party, European Comm'n, Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679, at 6 (last revised and adopted Oct. 4, 2017), available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

– DeepMind, https://deepmind.com/research/case-studies/alphago-the-story-so-far (detailing the story of AlphaGo, the first computer program to defeat a professional human Go player)

– Direction de l'action du Gouvernement, publications officielles et information administrative, Sénat 146 et 147 (2018-2019)

– Directive on Automated Decision-Making, Gov't of Canada (Feb. 5, 2019), https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

– Employment Tests and Selection Procedures, U.S. Equal Emp't Opportunity Commission, https://www.eeoc.gov/laws/guidance/employment-tests-and-selection-procedures (last visited July 5, 2020)

– Ethics Guidelines For Trustworthy AI, Eur. Commission (Apr. 8, 2019), https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

– European Union Agency for Fundamental Rights, #BigData: Discrimination In Data-Supported Decision-making, FRA (2018), https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf

– FTC Website, https://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy-security

– Gender Equality Seal for Public and Private Organizations, UNDP, https://www.undp.org/content/undp/en/home/2030-agenda-for-sustainable-development/people/gender-equality/transforming-workplaces-to-advance-gender-equality/gender-equality-seal-for-public-and-private-enterprises.html (last visited July 5, 2020)

– GLOBAL RESEARCH COUNCIL, SUPPORTING WOMEN IN RESEARCH: POLICIES, PROGRAMS AND INITIATIVES UNDERTAKEN BY PUBLIC RESEARCH FUNDING AGENCIES (Apr. 2019), https://anr.fr/fileadmin/documents/2019/GRC_GWG_Case_studies_final.pdf

– Global Research Council, Supporting Women in Research: Policies, Programs and Initiatives Undertaken by Public Research Funding Agencies, (2019), https://anr.fr/fileadmin/documents/2019/GRC_GWG_Case_studies_final.pdf.

– Guidelines on Automated Individual Decision-making and Profiling for the Purposes of Regulation 2016/679, Eur. Commission 10 (2017), available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

– House of Commons and Women & Equalities Comm., Enforcing the Equality Act: the law and the role of the Equality and Human Rights Commission–Tenth Report of Session 2017–19, House of Commons (2019), https://publications.parliament.uk/pa/cm201719/cmselect/cmwomeq/1470/1470.pdf

– Human Rights Impact Assessments, Nat'l Action Plans on Bus. & Human Rights, https://globalnaps.org/issue/human-rights-impact-assessments/ (last visited July 5, 2020)

– Jobscan Blog (Nov. 7, 2019), https://www.jobscan.co/blog/99-percent-fortune-500-ats/

– Legislation, N.Y.C. City Council, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search (click Details tab) (last visited July 5, 2020)

– Letter from Kamala D. Harris, U.S. Senator, et al. to Victoria A. Lipnic, Acting Chair, U.S. Equal Opportunity Emp't Comm'n, et al. (Sept. 17, 2018), available at https://www.hrpolicy.org/downloads/2018/Sen%20Harris%20Ltr%20to%20EEOC%20on%20Facial%20Recognition%20Software.pdf

– National Conference of State Legislatures, Sex and Gender Discrimination in the Workplace, https://www.ncsl.org/research/labor-and-employment/-gender-and-sex-discrimination.aspx

– OECD AI Policy Observatory: A Platform to Share and Shape AI Policies, Official Launch Agenda at 1 (Feb. 27, 2020), https://www.oecd.org/going-digital/ai/oecd-ai-observatory-launch-agenda.pdf

– Press Release, IEEE Standards Ass'n, IEEE Invites Companies, Governments and Other Stakeholders Globally to Expand on Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) Work (Feb. 26, 2020), https://standards.ieee.org/news/2020/phase-1-ecpais.html

– Press Release, Oracle, New Study: 64% of People Trust a Robot More Than Their Manager (Oct. 15, 2019), https://www.oracle.com/corporate/pressrelease/robots-at-work-101519.html
Pymetrics, https://www.pymetrics.com/employers (last visited July 5, 2020)

– Qualified Statewide Ballot Measures, Proposition 24, SECRETARY OF STATE OF CALIFORNIA, https://www.sos.ca.gov/elections/ballot-measures/qualified-ballot-measures (last visited September 15, 2020)

– Reforming The U.S. Approach to Data Protection and Privacy, Council on Foreign Rel. (Jan. 30, 2018), https://www.cfr.org/report/reforming-us-approach-data-protection

– Sci. Foresight Unit (STOA), European Parliamentary Research Serv., A Governance Framework for Algorithmic Accountability and Transparency, Eur. Parliament 55 (Apr. 2019), http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/

– EPRS_STU(2019)624262_EN.pdf

– Sci. Foresight Unit (STOA), European Parliamentary Research Serv., Understanding Algorithmic Decision-making: Opportunities and Challenges, Eur. Parliament 4 (Mar. 2019), http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261(ANN1)_EN.pdf

– Some Basic Concepts, Information Comm'r's Office (U.K.), https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-data-protection/some-basic-concepts/ (last visited July 5, 2020)

– The Age of Digital Interdependence: Report of the UN Secretary-General's High-level Panel on Digital Cooperation 4 (2019), https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf

– The British Psychological Society, Psychological testing: A test user's guide (2017), https://ptc.bps.org.uk/sites/ptc.bps.org.uk/files/guidance_documents/ptc02_test_users_guide_2017_web.pdf

– U.S. Fed. Trade Comm'n, Protecting Consumer Privacy in an Era of Rapid Change Recommendations for Business and Policymakers 12-13 (Mar. 2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf

– UN News, Generations of progress for women and girls could be lost to COVID pandemic, UN chief warns, UN News, Aug. 31, 2020, https://news.un.org/en/story/2020/08/1071372

– United Nations Educational, Scientific and Cultural Organization (UNESCO), I'd Blush If I Could: Closing Gender Divides In Digital Skills Through Education, 19 (2019), https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

– United Nations, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, U.N. Doc. No. HR/PUB/11/04 (2011), https://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf

– United Nations, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, U.N. Doc. No. HR/PUB/11/04 (2011), https://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf

– What Does The GDPR Say About Automated Decision-making and Profiling?, Information Comm'r's Office (u.k.), https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-does-the-gdpr-say-about-automated-decision-making-and-profiling/ (last visited June 8, 2020)

– White Paper On Artificial Intelligence—A European Approach To Excellence And Trust, Eur. Commission (Feb. 19, 2020), https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

– Who Does The Data Protection Law Apply To?, Eur. Commission, https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/application-regulation/who-does-data-protection-law-apply_en (last visited June 8, 2020)

### Cases

– Cour de Cassation [Cass.], Supreme Court for Judicial Matters, soc. June 6, 2012, Bull. civ. V, No. 168

– United Steelworkers of America v. Weber, 443 U.S. 193, 202-03 (1979)

– Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 987 (1988)

### Constitutional Provisions

– 1958 Const. (Fr) art. 1

### Statutes

– Algorithmic Accountability Act of 2019, Congress, https://www.congress.gov/bill/116th-congress/senate-bill/1108/committees (last visited July 5, 2020)

– Artificial Intelligence Video Interview Act, 820 ILCS 42/1-15 (2019)

– Artificial Intelligence Video Interview Act, Ill. Gen. Assembly HB2557 (Aug. 9, 2019), http://www.ilga.gov/legislation/101/HB/PDF/10100HB2557lv.pdf

– California Consumer Privacy Act of 2018, Cal. Civ. Code §§ 1798.100 to 1798.199

– Children's Online Privacy Protection Act (COPPA), 15 U.S.C. §§ 6501-6505 (1998)

– Civil Rights Act of 1964, 42 U.S.C. § 2000e et seq (1964)

– Code du travail [C. trav.] [Labor Code]

– Code penal [C. pén.] [Criminal Code]

– Council Directive 2004/113/EC, 2004 O.J. (L 373)

– Council Directive 2006/54/EC, 2006 O.J. (L 204)

– Council Directive 2011/83/EU, 2011 O.J. (L 304)

– Dodd-Frank Act § 165(d)(1), 12 U.S.C. § 5365(d)(1)

– Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, 124 Stat. 1376, 1871 (2010)

– Equality Act 2010, c. 15, § (4) (U.K.)

– Fair Credit Reporting Act (FCRA) 15 U.S.C. § 1681 (1992)

– H.B. 1202, 441st Gen. Assemb., Reg. Sess. (Md. 2020)

– H.B. 2991, 101st Gen. Assemb. (Ill. 2019); H.B. 3415, 101st Gen. Assemb. (Ill. 2019)

– H.R. 2231, 116th Cong. (2019); S. 1108, 116th Cong. (2019)

– Health Insurance Portability and Accountability Act (HIPPA), Pub. L. No. 104-191, 100 Stat. 2548 (codified as amended in scattered sections) (1996)

– New York City, N.Y., A Local Law to amend the administrative code of the city of New York, in relation to the sale of automated employment decision tools, Int 1894-2020 (2020)

– The Liability for Defective Products Directive (1999/34/EC)

– The Product Liability Directive 85/374/EEC

WOMEN AT THE TABLE