Fabian Lütz

Discrimination by Correlation

Towards Eliminating Algorithmic Biases and Achieving Gender Equality

> "Accountability requires human judgement, and only humans can perform the critical function of making sure that, as our social relations become ever more automated, domination and discrimination aren't built invisibly into their code"[1] (Frank Pasquale).

The analysis[2] focuses on opportunities and challenges of algorithms[3] and risks in the "algorithmic age"[4] and will explore avenues to address the impact of algorithms[5] in the area of gender equality (GE) law regarding biases and discrimination.

I. Obey and Disobey—the Terms Imposed by Behavior Changing Algorithms and Gender-based Discrimination

In most online activities[6] consumers' human intelligence[7] is confronted with decisions of algorithms. Consumers have to obey or dis-obey. Often there is no real choice. Not accepting the terms and conditions imposed by companies equals exclusion from the service, which can be best described by the term of *behavior changing algorithms*[8]. Some platforms face no competition, exercise monopoly[9] or "algorithmic power" and could be viewed as

1 Pasquale, Frank: The Black Box Society: The Hidden Algorithms Behind Money and Information, Boston 2015, p. 213.

2 The author would like to thank Tim Papenfuss for practical insights and comments on an early draft.

3 See Russell, Stuart: Artificial intelligence: The future is superintelligent. In: Nature 548 (2017), p. 520–521. ; Bostrom, Nick: Superintelligence. Paris 2017; Bostrom, Nick: The future of humanity. In: Geopolitics, History, and International Relations (2009), 1(2), 41–78. Tegmark, Max: Life 3.0: Being human in the age of artificial intelligence, London 2017.

4 Louridas, Panos: Algorithms, Boston 2020, p. 1.

5 For a critical perspective, see Gunkel, David J.: The Machine Question: Critical Perspectives on AI, Robots, and Ethics, Cambridge 2012.

6 Search, applications for job postings or unemployment benefits, online advertisements (ads) for products or recommendations for books.

7 Badre, David: On Task, Princeton 2020, p. 47.

8 This terminology is inspired by Zuboff, Shoshana: The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power, New York 2019.

9 See in general, Petit, Nicolas: Big Tech and the Digital Economy: The Moligopoly Scenario. Oxford 2020.

gate keepers[10]. In a democracy nobody should be discriminated because of gender when using services[11]. But what are algorithms? Barocas defines an algorithm as "a formally specified sequence of logical operations that provides step-by-step instructions for computers to act on data and thus automate decisions".[12]. Algorithms understood as a list of step-by-step instructions which are nourished with real world data, have an objective and follow the instructions or mathematical operations to achieve the defined aim[13]. Fry groups algorithms into four main categories according to the tasks: 1) priorization[14], 2) classification[15], 3) association[16] and 4) filtering[17]. These algorithms can come in the shape of either "rule-based algorithms" where instructions are programmed by a human or "machine-learning algorithms"[18]. The article will mostly refer to algorithms in general[19].

10 See Article 3 (1) of the Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act).

11 See Zuiderveen Borgesius, Frederik: Strengthening legal protection against discrimination by algorithms and AI, In: The International Journal of Human Rights, 24:10 (2020), p. 1572–1593, highlighting the threat of AI to the right to non-discrimination.

12 Barocas, Solon: Data & Civil Rights: Technology Primer (2014); In essence, algorithms are "a step-by-step procedure for solving a problem or accomplishing some end", see Algorithm." Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/algorithm (February 7, 2021).

13 See Fry, Hannah: Hello World: How to be Human in the Age of the Machine, London 2018, p. 8–9.

14 One classic example is search engines that rank different results or online video platforms suggesting what movies to watch, ibid., p. 9.

15 An example is online advertisement by showing different categories of people different advertisements, ibid., p. 10.

16 This task is important for this analysis, as it tries to find relationships, connections and correlations between things, used for example by online book stores to make recommendations, ibid., p. 10.

17 This task removes noise from signals by filtering information, a type of task used by speech recognition or social media applications, ibid., p. 10–11.

18 Ibid., p. 11–12: "You give the machine data, a goal and feedback when it's on the right track—and leave it to work out the best way of achieving the end", ibid., p. 12.

19 Wooldridge 2020, p. 349. Under the umbrella term of narrow artificial intelligence, machine learning (ML) is a sub-category and (artificial) neural networks or deep learning further sub-categories. Boden is classifying 5 different forms of AI: symbolic artificial intelligence, artificial neural networks, evolutionary programming, cellular automata and dynamical systems, see Boden 2010, p. 6. For an introduction to algorithms see Louridas 2020, p. 181f; for an introduction to Deep Learning and the relationship between algorithm, machine learning and deep learning, see Keller, John D: Deep

Obey shall be understood in two ways: first, humans must obey the terms imposed by companies to use systems and second, the state can impose regulation on companies they need to obey to. Dis-obey shall be understood as humans dis-obeying in order to preserve their rights[20], notably in the absence of legal rules or if companies dis-obey regulatory attempts to preserve their business model. The *dis-obey* approach could inspire consumers to follow *rights-preserving behavior,* such as *data poor* approaches, favoring data friendly companies, introducing "noise" into their data supply or avoid digital services that potentially discriminate[21]. Considering this tension between *obey* and *dis-obey,* regulators have been reflecting on rules for fair and non-discriminatory algorithms. The European Commission (EC) published a draft Regulation (Artificial Intelligence Act)[22] on 21 April 2021, following the adoption of the Digital Services Act (DSA)[23] and the Digital Markets Act (DMA)[24]. Many international bodies have adopted standards on AI (OECD[25],

Learning, Boston 2019, p. 6. ML can be subdivided into supervised and un-supervised learning.

20 For example, by choosing alternative ways of using services offered by companies.

21 Consumers could use algorithms to detect discriminatory algorithms.

22 European Commission, Proposal for a Regulation of the European Parliament and the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM (2021) 206 final.

23 The DSA tries to mitigate some of the risks for women: "Specific groups [...] may be vulnerable or disadvantaged in their use of online services because of their gender [...] They can be disproportionately affected by restrictions [...] following from (unconscious or conscious) biases potentially embedded in the notification systems by users and third parties, as well as replicated in automated content moderation tools used by platforms."

24 The European Digital Strategy consist of the Digital Services Act (DSA) and the Digital Markets Act (DMA):Proposal for a Regulation of the European Parliament and the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM/2020/825 final and Proposal for a Regulation of the European Parliament and the Council on contestable and fair markets in the digital sector (Digital Markets Act), COM/2020/842 final.

25 OECD, Principles on Artificial Intelligence, https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm

Council of Europe[26] or UNESCO[27]). The EC's Advisory Committee on Equal Opportunities for Women and Men adopted an opinion on AI and GE[28], containing recommendations to address algorithmic biases and prevent gender-based discrimination.

A case of discrimination usually concerns individual cases, but the impact can reach societal scale when patterns of algorithmic discrimination evolve and reinforce biases and discrimination[29]. Each discriminated individual will be reflected in the datasets and contribute to create future risks of discrimination for women and men as categorized and classified by algorithms. However, humans also rely on automatic processing of data by schematizing and grouping people in boxes, for example by sex or race[30]. Such classification and generalization could base decisions on a group of women or men to the detriment of an individual, which impacts the well-being of consumers using products and services[31] that rely on technology or workers accessing the labor market[32]. Moreover, one of the problems is the opaque decision making of algorithms, or "black box" as used by Pasquale to describe the

26 CoE Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, 8th April 2020, https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154 "ensure that racial, gender and other societal and labour force imbalances that have not yet been eliminated from our societies are not deliberately or accidentally perpetuated through algorithmic systems, as well as the desirability of addressing these imbalances through using appropriate technologies" (Preamble).

27 UNESCO, Report on AI and Gender Equality, https://unesdoc.unesco.org/ark:/48223/pf0000374174 (February 6, 2021).

28 European Commission, Advisory Committee on Equal Opportunities for Women and Men, Opinion on Artificial Intelligence (2020), https://ec.europa.eu/info/sites/info/files/aid_development_cooperation_fundamental_rights/opinion_artificial_intelligence_gender_equality_2020_en.pdf (February 6, 2021).

29 In general, see Adam, Alison: Artificial knowing: gender and the thinking machine. London 1998.

30 Kleinberg, Jon; Ludwig, Jens; Mullainathan, Sendhil; Sunstein, Cass R.: Algorithms as discrimination detectors. In: Proceedings of the National Academy of Sciences Dec 2020, 117 (48), p. 30097.

31 Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services.

32 Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast).

fact that the inner workings of an algorithm are sometimes difficult to grasp, especially for potential victims of discrimination.

Relying on literature and current institutional proposals, the article assesses the opportunities and risks both for regulating and using algorithms. Dealing with the topic of AI and gender from the angles of "regulatory object" and "useful tool" will shed new light and contribute to an ethical[33] and fair framework[34] to enforce GE laws.

II. From Classical Discrimination towards Discrimination by Correlation

Before explaining discrimination by correlation (3) and giving examples (4), I will present the relevant EU law and discuss the concept of discrimination (1) as well as the relationship between algorithms and bias (2).

1) Some Reflections on EU Law and Gender-based Discrimination

EU anti-discrimination law works with the concept of protected characteristics (Ex. gender or age). However, this becomes increasingly difficult when decisive elements in the decision-making result not from humans but algorithms. Current laws were adopted before the age of algorithms and are not equipped to deal with all new legal challenges even if formulated in an abstract and general way to deal with (un)foreseen situations[35]. Judges will have to interpret existing laws in light of technological developments, which could accommodate AI. EU law distinguishes between direct and indirect discrimination. A direct discrimination in EU law[36] exists "where one person

33 Liao, S. Matthew: A Short Introduction to the Ethics of Artificial Intelligence. Ethics of Artificial Intelligence. Oxford 2020.

34 This article will not discuss fair and ethical AI in general, see notably Coeckelbergh 2020.

35 On the nature of the abstract general design and force of the law, see Hart, Herbert Lionel Adolphus; Green, Leslie: The concept of law. Oxford 2015, p. 21.

36 US law differentiates along the lines of disparate treatment and disparate impact, therefore choosing a similar classification but closer to the dichotomy known under competition law as "by object/by effects approach", see for U.S. law the exhaustive overview by Barocas, Solon; Selbst, Andrew D.: Big Data's Disparate Impact. In: California Law Review 104 (2016), p. 671–732.

is treated less favorably on grounds of sex than another is, has been or would be treated in a comparable situation"[37]. Indirect discrimination "where an apparently neutral provision, criterion or practice would put persons of one sex at a particular disadvantage compared with persons of the other sex, unless that provision, criterion or practice is objectively justified by a legitimate aim, and the means of achieving that aim are appropriate and necessary"[38]. While direct discrimination cannot be justified in principle, a possibility for justification exists for indirect discrimination. A different treatment is not discriminatory, when it is justified, appropriate and necessary (proportionality test)[39]. Procedurally, the burden of proof is essential in non-discrimination cases because the claim for a discrimination needs to be supported by evidence, which is generally shared between the victim and the "alleged" discriminator[40]. Once a *prima facie* evidence is brought by the victim, the "discriminator" needs to rebut the claim, a process called the shifting of the burden of proof. The idea is to facilitate the access to evidence for the claimant, often difficult, especially in cases involving opaque algorithmic decision procedures. In the case *Schuch–Ghannadan*[41], the Court of Justice of the European Union (CJEU) refined its jurisprudence by ruling that the burden of proof does not require bringing statistical data or facts (beyond some prima facie evidence), if the claimant has no or difficult access[42]. This jurisprudence *de facto* extends the rights of victims of discrimination, defining what is expected of them in terms of evidence. Concretely, only evidence that is not more than reasonable to access can be expected which is of relevance for cases involving AI. This case law could facilitate bringing claims against companies. If the claimant cannot reasonably access the information contained in the algorithm, the "burden of proof" shifts to the company which needs to

37 Article 2 (1)(a) Directive 2006/54/EC.

38 Article 2 (1)(b) Directive 2006/54/EC.

39 See Craig, Paul; Gráinne De Búrca: EU law: text, cases, and materials. Oxford 2020, p. 544–545.

40 See for the discrimination test, Ellis, Evelyn; Watson, Philippa: EU anti-discrimination law, Oxford 2012.

41 C-274/18, Minoo Schuch-Ghannadan v Medizinische Universität Wien, EU:C:2019:828.

42 Ibid: "Art. 19 Abs. 1 der Richtlinie 2006/54 ist dahin auszulegen, dass er von der Partei, die sich durch eine solche Diskriminierung für beschwert hält, nicht verlangt, dass sie, um den Anschein einer Diskriminierung glaubhaft zu machen, in Bezug auf die Arbeitnehmer, die von der nationalen Regelung betroffen sind, konkrete statistische Zahlen oder konkrete Tatsachen vorbringt, wenn sie zu solchen Zahlen oder Tatsachen keinen oder nur schwer Zugang hat."

show that the algorithm did not discriminate which incentivizes companies to avoid discrimination in the first place. Previously, the CJEU was reluctant in terms of access to information when it excluded in *Meister*[43] a "right [...] to have access to information indicating whether the employer has recruited another applicant" even when the job applicant "claims plausibly that he meets the requirements listed in a job advertisement and whose application was rejected"[44]. This represented an obstacle for job applicants that were refused by algorithms to get access to the underlying data that influenced the decision outcome, making proof of algorithmic discrimination more difficult than classic discrimination. The CJEU had no opportunity (yet) to clarify its interpretation in a case of AI[45], but would dispose of tools to facilitate confidential access to data, for example via in *camera* procedures to protect business secrets or consulting AI experts to give expert evidence.

Statistical analysis is used for risk assessment by insurance companies to deal with complexity, sometimes to the detriment of accuracy. Insurance companies used *gender* to distinguish between different risks, to establish price differentiation by *gender* in car insurance contracts[46]. The case "Test-Achats" concerned the practice of using *gender* for insurance premiums[47]. The CJEU ruled that considering *gender* for calculating insurance premiums is discriminatory, obliging the firms to introduce *gender* neutral insurance contracts. Despite not being directly linked to AI, the case gives guidance to assess potential discriminations for situations of statistical data and data sets used by algorithms where a similar process of generalization exists. Even if the CJEU "banned "using *gender*-specific insurance contracts, algorithms can easily circumvent this prohibition by using criteria or so-called proxies, to infer the *gender* of a person. Consequently, it remains to be seen how courts would decide a case involving algorithms and if the concept of discrimination is still well equipped to "grasp" the essence of algorithmic

43 CJEU, C-415/10 Galina Meister v Speech Design Carrier Systems GmbH EU:C:2012:217.

44 Ibid, para. 49.

45 Some guidance was received from the CJEU in Seymour-Smith, that "mere generalizations concerning the capacity of a specific measure to encourage recruitment are not enough to show that the aim of the disputed rule is unrelated to any discrimination based on sex nor to provide evidence on the basis of which it could reasonably be considered that the means chosen were suitable for achieving that aim." CJEU, Case C-167/97, Seymour-Smith, EU:C:1999:60.

46 CJEU, C-236/09.

47 Ibid.

discrimination on the basis of *gender*[48] ? The borders of the protected characteristics such as gender or race become increasingly blurred when algorithms are involved. Algorithms might replace the distinguishing element of *gender* by other data via correlation. One can expect more *discrimination by correlation* based on datasets that correlate and infer information indirectly and discriminate in fine by gender. I call this process *discrimination by correlation,* which is not restricted to AI. The detection of hidden or indirect mechanisms should interest the regulator, as traditional discrimination patterns (e.g. *gender* or *age*) risk becoming less frequent. Algorithms could play a role as "discrimination detectors" [49] for the regulator.

2) Towards Discrimination by Correlation: Algorithms Reflecting and Magnifying Gender Biases?

The assumption "machine learning is fair by default"[50] is disputable, as algorithms can "potentially increase bias and discrimination"[51]. Algorithms are seen as "neutral", as they simply "execute code" based on available data. However, algorithms are only as neutral as the datasets. The outcome of an algorithm could be amplifying biases because of gender-biased data (see further in II. 1 on the problem of the *gender data* gap)[52].

Recent literature is assessing potential impacts of algorithms on gender-based discrimination[53]. Bias is no new phenomena; it has been known

48 On EU law, see Xenidis, Raphaële; Senden, Linda: EU Non Discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination. In: Bernitz, Ulf; Groussot, Xavier; de Vries, Sybe A. (Eds.), General Principles of EU law and the EU Digital Order, Bruxelles 2020, pp. 151–182.

49 Kleinberg et. al. 2020, p. 30096.

50 Argued by Geerts, Thierry: Homo Digitalis, Lanno 2021; Hardt, Moritz. In: Medium 2014, How big data is unfair, https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de (February 6, 2021).

51 Coeckelbergh 2020, p. 75.

52 This risk is also highlighted in the standard text book on AI: "Often, the data themselves reflect pervasive bias in society", see Russell, Stuart; Norvig, Peter: Artificial intelligence: a modern approach, London 2021, p. 49.

53 See the special report European Commission, "Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law" (2021) prepared by the Legal Network of Gender Experts of the EC, https://www.equalitylaw.eu/downloads/5361-algorithmic-discrimination-in-europe-pdf-1-975

by enforcers analyzing discriminatory behavior[54]. But biases in algorithmic discrimination leverage and amplify discriminatory risks and lead to new forms of discrimination such as invisible discrimination. Discriminations involving algorithms are more complex and require refined detection mechanisms. As the decision-making process inside the algorithm is often unclear, discriminations might occur without ever being detected (or consciously felt), due to the opaque nature of the AI. Job ad recommendations not shown by the algorithm on the screen because women have not been defined as target audience, would be unthinkable in the real world. A billboard at the side of the highway or next to the bus stop would not "discriminate" on the basis of *gender* of the viewer.

Cause and origin of discrimination is not necessarily a protected characteristic (Ex. *gender*). If certain job postings are not shown to women because of the characteristic gender, algorithms decide/learn and recognize patterns based on available data. Data and the correlation between data points enable the algorithms to conclude that specific ads should not be shown to women. As algorithms correlate information from datasets on which they have been trained, *discrimination by correlation* grasps this new reality of algorithmic discrimination as it describes how discrimination occurs: by correlating data, without being able to identify which specific data points have caused a decision that is discriminatory.

3) Examples of Gender Bias and Discrimination by Correlation

Four examples illustrate the reflections in the areas of (a) online ads, (b) employment, (c) image processing and (d) natural language processing, where biases/stereotypes or (gender-based) discrimination occurs. There is increasing awareness about discrimination and inequalities occurring in online platforms[55], when it comes to determining recidivism for criminal

54 Coeckelbergh 2020, p. 125.

55 Renan Barzilay, Arianne: The Technologies of Discrimination: How Platforms Cultivate Gender Inequality. In: The Law & Ethics of Human Rights 13 (2019), no. 2, p. 179–202.

convicts[56] or predicting the likelihood of a future crime[57]. The predictive power of algorithms used by supermarkets to "predict" pregnancy[58] based on the products purchased[59] gained media attention. Feminist literature [60] and recent books expose that the data-driven world and algorithms are often designed by and for men[61]

a) Online Advertisements

Google and *Facebook* show targeted ads to users using algorithms [62]. Experimental research by Lambrecht/Tucker[63] revealed that women received less job ads for STEM[64] professions than men. The authors explored how algorithms deliver gender neutral job ads promoting job opportunities in the STEM sector. Despite gender neutrality, empirical evidence revealed that fewer women saw the ad despite a similar estimation of "click-through-rate". This is explained by the so-called "Gender Valuation Gap"[65] which

56 Skeem, Jennifer; John Monahan; Christopher Lowenkamp: Gender, risk assessment, and sanctioning: The cost of treating women like men. In: Law and human behavior 40.5 (2016), p. 580; Wright, Emily M.; Salisbury, Emily J.; Van Voorhis, Patricia: Predicting the prison misconducts of women offenders: The importance of gender-responsive needs. In: Journal of Contemporary Criminal Justice 23.4 (2007), p. 310–340; DeMichele, Matthew; Baumgartner, Peter; Wenger, Michael; Barrick, Kelle; Comfort, Megan; Misra, Shilpi: The Public Safety Assessment: A re-validation and assessment of predictive utility and differential prediction by race and gender in Kentucky (2018).

57 Mayson, Sandra G.: Bias in, bias out. In: Yale Law Journal 128 (2018), p. 2218.

58 Zuiderveen 2018, p. 13 as example for intentional discrimination based on gender which would be difficult to prove.

59 Basdevant, Adrien; Mignard, Jean-Pierre: L'Empire des données. Essai sur la société, les algorithmes et la loi. Paris 2018, p. 91f.

60 Wellner, Galit; Rothman, Tiran: Feminist AI: Can We Expect Our AI Systems to Become Feminist? In: Philosophy & Technology. 33 (2020), p. 191–205.

61 Perez, Caroline Criado: Invisible women: Exposing data bias in a world designed for men. London 2019.

62 Agrawal, Ajay; Joshua Gans; Avi Goldfarb: Máquinas predictivas: la sencilla economía de la inteligencia artificial, Madrid 2019, p. 238–241.

63 Lambrecht, Anja; Tucker, Catherine E.: Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads. In: Management Science 65 (2019), p. 2966–2981.

64 Science Technology Engineering and Mathematics.

65 The 21% Gender Valuation Gap, https://www.wordstream.com/blog/ws/2014/05/13/gender-bias (February 6, 2021).

means ads are more expensive to show to women, as "women [are] being undervalued by 21% in online marketing"[66], which refers to the potential value per click and earnings from ads. As algorithms are run cost effectively, the companies prefer to show ads to men even for gender-neutral ads. This could represent a discriminatory risk if women are systematically excluded from seeing the ads[67]. Research revealed [68] the potential unequal treatment for men and women in image recognition algorithms for advertising[69], when inserting gender stereotypes into the datasets. Researchers concluded that *Facebook* could determine precisely to whom ads are targeted, which shows the discriminatory potential[70]. Referring to Lambrecht/Tucker, computer scientists developed a commendable strategy[71] to achieve fairer ads without gender bias[72]. Research and specific algorithms hint at the possibility to control discrimination in online advertisement auctions[73].

66 Criado et al. 2020, p. 1.

67 Lambrecht; Tucker 2018.

68 See the examples by Orwat, Carsten: Risks of Discrimination through the Use of Algorithms. A study compiled with a grant from the Federal Anti-Discrimination Agency. Berlin 2020, p. 37.

69 Ali, Muhammad; Sapiezynski, Piotr; Bogen, Miranda; Korolova, Aleksandra; Mislove, Alan; Rieke, Aaron (2019): Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes. In: arXiv e-prints, https://arxiv.org/pdf/1904.02095.pdf (February 6, 2021).

70 The images were only readable for the algorithm but not for humans and targeted either men or women depending on the stereotype "coded" into the pictures beforehand.

71 Methodology, http://cs.yale.edu/bias/blog/jekyll/update/2019/02/08/fair-advertising.html (February 6, 2021).

72 Demo, https://fair-online-advertising.herokuapp.com (February 6, 2021).

73 Celis, L. Elisa ; Mehrotra, Anay; Vishnoi, Nisheeth: Toward controlling discrimination in online ad auctions. In: International Conference on Machine Learning. PMLR, 2019.

b) Employment and Recruitment

Algorithms are used at all stages of employment[74]. One example is Amazon's recruitment algorithm which discriminated women[75]. They are also used for the distribution of unemployment benefits which potentially discriminated women, where algorithms classify unemployed people into three categories in accordance with job prospects, therefore a classical exercise of sorting. In *concreto*, women received different scores than men, notably due to absences in the labor market (maternity and parental leave)[76]. This algorithm is problematic[77] because gender, labor market absences, births and family leaves are incorporated into the predictions[78] of job prospects. Private companies and national administrations use algorithms to guide and (improve?) decision-making. Even if in the Austrian example, the court rejected claims of discrimination by the algorithm, legal scholars and computer scientists criticized this algorithm for using criteria that are strongly linked or associated with one sex. Using maternity leave, family leave (dominantly taken by women) or military service (mostly men) besides the protected characteristic of *gender* is problematic. Discrimination can be difficult to detect and is sometimes easily confused in some situations. Known as the "Simpson's paradox", statistics do not necessarily reveal the underlying reasons for different (potentially discriminatory) outcomes reflected in a

74 Pre-employment, recruitment, employment including promotions and evaluations, post-employment, unemployment benefits.

75 Reuters, 11th October 2018, Amazon scraps secret artificial intelligence recruiting tool that showed bias against women, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (February 6, 2021).

76 https://verfassungsblog.de/koennen-algorithmen-diskriminieren/ (February 6, 2021).

77 For a schematic overview of some of the parameters used, see https://www.derstandard.at/story/2000089720308/leseanleitung-zum-ams-algorithmus ; https://algorithmwatch.org/en/story/austrias-employment-agency-ams-rolls-out-discriminatory-algorithm/ and here for a critical assessment of the Austrian Academy of Sciences that discussed potential discriminatory effects of the algorithm https://www.oeaw.ac.at/ita/projekte/2020/der-ams-algorithmus. The algorithm was stopped temporarily in 2020, and then again authorized by court decision in December 2020: https://netzpolitik.org/2020/automatisierte-entscheidungen-gericht-macht-weg-fuer-den-ams-algorithmus-wieder-frei/

78 For predictions made by algorithms, Spiegelhalter, David: The art of statistics: learning from data, London 2019, p. 143.

statistic, which was revealed in the context of university admission[79]. Such arguments can easily be used by defendants in court to refute alleged cases of discrimination[80].

In general, an International Labor Organization (ILO) report sums up the challenges of AI used in employment: "[...] An automated recruitment system based on analyzing historic data would replicate [...] bias, thereby reinforcing pre-existing discrimination"[81]. An example for discriminatory treatment at work,[82] is "classification bias", which "occurs when employers rely on classification schemes, such as data algorithms, to sort or score workers in ways that worsen inequality or disadvantage along the lines of [...] sex, or other protected characteristics"[83]. Finally, Kleinberg et al. describe a hypothetical example of alleged discrimination where a tech company is not hiring a woman. They compare a human decision to discriminate with a potential AI-based decision to discriminate[84]. For classification, attempts have been made to achieve fair and non-discriminatory outcomes of the algorithm[85].

79 Ibid, p. 110–112.

80 Kleinberg et al. 2020, p. 30097: "Challenges in using statistical evidence to show intentional discrimination, small sample sizes, unclear objectives, and the general opacity of human cognition combine to create a fog of ambiguity, which prevents us from stopping a behavior that we know to be widespread yet for which in any one instance there may well be plausible alternative explanations."

81 Ernst, Ekkehardt; Merola, Rossana; Samaan, Daniel: Economics of artificial intelligence: Implications for the future of work. In: IZA Journal of Labor Policy 9.1 (2019), p. 16.

82 Kim, Pauline T.: Data-driven discrimination at work. In: William & Mary Law Review 58 (2016), p. 857–866.

83 Ibid., p. 866.

84 Kleinberg et al. 2020, p. 30096–30097.

85 Concrete examples for code to be used to avoid discriminatory outcomes on the basis of sex: https://github.com/Trusted-AI/AIF360/blob/master/examples/demo_meta_classifier.ipynb (February 6, 2021).

c) Image Processing

The photo tagging algorithm of Google created a discriminatory concern, when black people were labeled erroneously as "gorillas"[86]. This relates to the problem described in section III.) on the (un)availability of data sourcing the algorithm as described by Hosaganar: "Image-processing algorithms that hadn't been trained on a large enough number of photos of black people were unable to account for different skin tones and lightning"[87]. As research on gender and racial biases highlights[88], similar observations were identified with regard to pictures of search engines. When searching for "CEO", much more pictures of men were displayed than women, but the percentage reflected was worse than the real ratio between men and women. An attempt to achieve a more balanced image search[89] has been developed by Celis, L. Elisa, et al.[90].

d) Natural Language Processing[91], Search and Autocomplete Functions

The auto-complete function is implemented in most search engines. Search queries feed the search algorithms, and this is fed back to suggest search terms while the user is typing the query. A useful tool without doubt, it could make suggestions (or predict the user's search intentions) in ways that do not necessarily match the real intention of the searcher, leading to a discriminatory behavior or reinforcing current discriminatory patterns. Examples reported by *The Economist* include a Dutch father searching for infor-

86 Hosanagar, Kartik: A human's guide to machine intelligence: how algorithms are shaping our lives and how we can stay in control, New York 2020, p. 44–45.

87 Ibid., p. 44–45.

88 Buolamwini, Joy; Timnit Gebru: Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Conference on fairness, accountability and transparency. PMLR (2018); Skeem 2016, p. 580.

89 See methodology, http://cs.yale.edu/bias/blog/jekyll/update/2018/01/20/balanced-news-search.html; see the demo: https://fair-image-search.herokuapp.com/image Diversity.php (February 6, 2021).

90 Celis, L. Elisa; Kapoor, Sayash; Salehi, Farnood; Vishnoi K. Nisheeth: An algorithmic framework to control bias in bandit-based personalization. In: arXiv:1802.08674(2018), https://arxiv.org/abs/1802.08674 (February 6, 2021).

91 For an overview of natural language processing, see Mitchell, Melanie: Artificial intelligence: A guide for thinking humans. London 2019, p. 223–251.

mation on parental leave and "how to combine work and fatherhood" with the auto-complete function suggesting: "When he searched for advice on combining fatherhood with work, the search engine asked if he had meant "motherhood and work". [92] This is not only discriminatory towards women (and men), it also perpetuates stereotypes and biases on gender roles and distorts reality. This might change over time[93] but the importance of search predictions without gender bias remains.

A second example revealed by Cadwalladr concerns ads highlighted by U.N. Women[94] which is also based on auto-complete suggestions[95]. According to the information, the ads revealed that if you type "women should" this leads to "women should stay at home" and "women should be slaves". Likewise, "women shouldn't" leads to "women shouldn't have rights" and "women shouldn't vote"[96]. Even if this reveals existing perceptions about society, stereotypes, and biases (as a result of people searching for this key words), there is a risk, that people are steered in a direction they would not have considered, creating a new audience for gender bias, stereotypes and attempts to discriminate. Some authors conclude that gender bias and racial bias enshrined in search engines like "Google's autocomplete is by no means an exception in the world of algorithms"[97].

Finally, another illuminating example[98] comes from the area of natural language processing (NLP), where it has been shown that word embeddings [99] can cause an amplification of existing bias, stereotypes and lead to discriminatory outcomes. Word embeddings are widely used in applications such as search or CV analysis. Some authors argue that the use of word embeddings is "blatantly sexist [...] and hence risk introducing biases of various

92 The Economist 7th October 2017, Men, women and work, https://www.economist.com/international/2017/10/07/the-gender-pay-gap (February 6, 2021).

93 The author repeated this search query on 10 October 2021.

94 https://www.unwomen.org/en/news/stories/2013/10/women-should-ads (February 6, 2021).

95 Hosanagar 2020, p. 42–43.

96 Ibid, p. 42.

97 Ibid, p. 44.

98 Mitchell 2019, p. 250–251.

99 Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James; Saligrama, Venkatesh, Kalaim Adam: Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In: arXiv:1607.06520 (2016).

types into real-world systems"[100]. Even gender-neutral wording is often associated as either female (homemaker, nurse, receptionist) or male (maestro, skipper, protégé). Whereas in analogy puzzles, "man is to king, as woman is to X", the best answer is "queen", however for other simple vector arithmetic, such word embeddings reveal potential sexism implicit in the words as shown in the following example: "man: woman :: computer programmer: homemaker". The authors of the empirical research propose a solution to this risk for biased data and have developed an algorithm detecting such risks in NLP[101]. NLP is vital for the advances in AI and is helpful to avoid biases and discrimination by correlation[102]. Another risk associated with NLP methods such as Word2vec[103] developed by Google is the de facto standard for neural networks to automatically learn word vectors. Programmers will have to "obey" this standard and use Google's database if they want to design a quality product. The influence of such neural networks is to "predict what words are likely to be paired with a given input word",[104] which is crucial for search. Considering the use of search in today's world and the occurrence of sexist and gender discriminatory outcomes caused by neural networks, the state needs to consider regulation. A possible solution to gender imbalance in rankings (for search, news feeds or recommendation systems), has been developed by computer scientists[105] in the framework of a Yale project "controlling bias in Artificial Intelligence"[106] including a demo version[107]. Such solutions could form part of the approach non-discrimination by code.

100 Bolukbasi et al., p. 11.

101 Ibid, p. 11; see also the strategy proposed by Ghili, Soheil; Ehsan Kazemi; Amin Karbasi: Eliminating latent discrimination: Train then mask. In: Proceedings of the AAAI Conference on Artificial Intelligence 2019, 33. № 01.

102 Mitchell 2019, p. 242.

103 https://code.google.com/archive/p/word2vec/ (February 6, 2021).

104 Mitchell 2019, p. 243.

105 http://cs.yale.edu/bias/blog/jekyll/update/2018/11/03/balanced-ranking.html (February 6, 2021).

106 http://balanced-ranking.herokuapp.com (February 6, 2021).

107 Ibid.

III. The (Un)available Data as Source for Gender Bias and Discrimination by Correlation

Algorithmic discrimination can be framed as a problem of how information processing by machines leads to gender-based discrimination. The AI White Paper of the EC highlights quite succinctly, that "without data, there is no AI. The functioning of [...] AI [...], and [its] actions and decisions [...] very much depend on the data set on which the systems have been trained. The necessary measures should therefore be taken to ensure that, where it comes to the data used to train AI systems, the EU's values and rules are respected, specifically in relation [...] the protection of fundamental rights."[108]

Even if the design of AI plays a role, in general, decisions or predictions are a direct result of the data[109]. Therefore, rather than focusing on the design stage (algorithmic processing bias)[110] and possible intentions of programmers[111], the present analysis will concentrate on the role of data as discriminations can occur regardless of whether the object of the algorithms is to discriminate or not.

108 European Commission 2020, White Paper AI, p. 19.

109 This was raised recently: "The question is where it is rooted—in the training dataset or in the algorithm?", see Wellner 2020; Strauß, S. From Big Data to Deep Learning: A Leap Towards Strong AI or 'Intelligentia Obscura'? In: Big Data Cogn. Comput. 2018, 2, p. 16.

110 See a cross-disciplinary perspective and a typology of three different biases relevant for the analysis of discrimination, Ferrer, Xavier, et al.: Bias and Discrimination in AI: a cross-disciplinary perspective. In: arXiv preprint arXiv:2008.07309 (2020), p. 1, https://arxiv.org/abs/2008.07309 (February 6, 2021).

111 However, diversity plays a decisive role, see Crawford, Kate; Whittaker, Meredith; Elish, Madelein Clare; Barocas, Solon; Plasek, Aaron; Ferryman, Kadija: The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term, New York 2016; European Commission, Gender Equality AI Opinion (2020), p. 8–9.

1) How History Defines the Future—the Problem of Gender Biased Data Sets

Data collection is limited to the available data and therefore algorithms also mirror bias (algorithmic bias[112])[113] present in society. The main source for biases[114] is data, in the form of input data and training data ("training bias"[115]). Consequently, the input data reflects the status of the world and societal perceptions. The algorithm identifies and finds data patterns during the training process and learns how to predict or advise decisions. If a gender bias is reflected in the data, it is likely to be incorporated in the algorithm. Thus, there is an increased risk for women of being discriminated if such patterns mirror stereotypes and biases.

The real danger for gender-based discrimination is that rather than identifying a protected characteristic (gender), it infers a person's protected characteristic based on available information. The power of algorithms is the result of correlating[116] data, making predictions based on (historic) data and by inference from non-existing characteristics. The algorithm could increase the risk of discrimination in apparently neutral situations where a characteristic such as gender is not known, explicitly excluded, or disregarded by the algorithm[117]. In essence, if there is no fully unbiased training data set, the algorithm will never be neutral.

112 Wooldridge 2020, p. 338.

113 For an overview of the 5 multiple sources of bias and discrimination (how target and class labels are defined; labelling the training data; collecting the training data; feature selection; and proxies), see Barocas; Selbst 2016, p. 677–693; Zuiderveen 2018, p. 10–13.

114 Next to the classification of Barocas; Selbst 2016, p. 677–693, some authors reduce the causes of algorithmic bias to two types ("biased training data and unequal ground truth"), see Hacker 2018, p. 5. The present analysis will take a different approach, focusing on the sources and entry points of bias around data.

115 See Ferrer et.al 2020, p. 1: "Algorithms learn to make decisions or predictions based on datasets that often contain past decisions. If a dataset used for training purposes reflects existing prejudices, algorithms will very likely learn to make the same biased decisions. Moreover, if the data does not correctly represent the characteristics of different populations, representing an unequal ground truth, it may result in biased algorithmic decisions."

116 A correlation means that there exists a relationship between facts, data or numbers and should not be confused with causation, see for example Spiegelthaler 2019, p. 96–99.

117 See notably, Barocas; Selbst 2016, p. 671.

The Council of Europe (CoE) Recommendation on the human rights impacts of algorithmic systems specifies on datasets: "In the design, development, [...] of algorithmic systems [...] States should carefully assess what human rights and non-discrimination rules may be affected as a result of the quality of data that are being put into and extracted from an algorithmic system, as these often contain bias and may stand in as a proxy for classifiers such as gender, race, [...]"[118]. The process of collecting data creates knowledge about consumers habits, referred to as profiling[119]. By profiling, companies are conducting a large-scale pattern recognition system that classifies consumers into categories. This facilitates decision-making by generalization and typification of consumers, such as recommending a product or showing a specific ad based on profiling[120]. Despite lacking accuracy, it is often a fast and cheap process for firms[121].

It has been argued by Hardt[122], that dominant groups tend to be favored by automated decision-making processes because more data is available and therefore receive fairer, more representative, and accurate decisions/predictions, than minority groups for which data sets are limited. The gender data gap captures this deficiency, as much less data on women is available in datasets[123]. Hardt even argues that accuracy could be considered as a proxy to fairness which means that women risk receiving fewer fair decisions by AI. Ensuring a fair data mining process could help de-bias and reduce discrimination. To sum up, AI "[...] raises difficult questions about how to ensure that

118 Recommendation CoE CM/Rec (2020)1 on the human rights impacts of algorithmic systems, point 2.2.

119 Hildebrandt, M. Profiling: From data to knowledge. In: DuD 30 (2006), p. 548–552.

120 See notably Anrig, Bernhard; Browne, Will; Gasson, Mark: The Role of Algorithms in Profiling. In: Hildebrandt, Mireille; Gutwirth, Serge (Eds.) Profiling the European Citizen. Berlin 2008, who distinguish two essential roles in data mining: "the procedure of the profiling process" and as a "mathematical procedure to identify trends, relationships and hidden patterns in disparate groups of data".

121 Packin, Nizan; Lev-Aretz, Yafit: Learning algorithms and discrimination. In: Research Handbook on the Law of Artificial Intelligence. Cheltenham, London 2018, p. 91, who highlight issues of reliability and data accuracy in the light of learning algorithms and discrimination.

122 Hardt 2014.

123 For the gender data gap, Kraft-Buchman, Caitlin; Arian, Réné: The Deadly Data Gap: Gender and Data. Geneva: Women at the Table (2019).

discriminatory effects resulting from automated decision processes, whether intended or not, can be detected, measured, and redressed".[124]

Having identified the source and nature of biases which could lead to discriminatory effects by algorithms, what to do about it?

2) Biases, Data Mining and Generalization

Data mining is the process of data collection to feed the algorithm. This is the first stage where biases and discrimination can be prevented. The type and quality of data collected is essential for creating fair and gender equal datasets as it influences potential and gravity of discrimination by the algorithm. Data quality, accuracy and representativeness are assets of good data sets. If datasets do not contain accurate, complete or any information on a specific group, it will be difficult to produce the desired (accurate) results of suggesting a behavior or predicting outcomes. Hence, collection and labelling of the data are crucial for training algorithms. Specific features, "the components of a piece of data that a ML program bases its decisions on"[125], need to be selected, in the stage during which a company filters and selects certain relevant criteria or data points. In this context, proxies are relevant to define the way the algorithm moves through the data and orients itself. Another challenge is the possibility for developers to hide anti-discriminatory behavior by masking the process. This prevents decision makers to (de)construct the data or manipulates the design of the algorithm by "hiding" and "covering" direct intentional discrimination[126].

Once data has been mined and datasets compiled, the algorithm uses generalization and concretization[127]. An authoritative book on stereotypes and probabilities has argued that "generalizations based on gender are important in their own right and as an illuminating beginning in considering

124 US White House, Executive Office of the President, Big Data: Seizing Opportunities, preserving, Washington 2014, http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report _5.1.14_final_print.pdf (February 6, 2021), point 64.

125 Wooldridge 2020, p. 343.

126 Zuiderveen 2018, p. 13.

127 Lee, Felicia R. : Discriminating? Yes. Discriminatory? № The New York Times, December 13, 2003. advance.lexis.com/api/document?collection=news&id=urn:contentItem:4B70-S120-01KN-20KW-00000-00&context=1516831 (February 6, 2021).

the circumstances under which using even statistically rational generalizations might be wrong"[128]. Problematic for non-discrimination by correlation is the so-called "tragedy of big data"[129], according to which the more variables or data you have, the more correlations that can show significance will be found by researchers or algorithms. According to Taleb, "falsity grows faster than information; it is nonlinear (convex) with respect to data".[130] and one problem associated with big data is that "there is a certain property of data: in large data sets, large deviations are vastly more attributable to noise (or variance) than to information (or signal)"[131]. This should caution regulators about the potential growing discriminatory nature of algorithms in ever increasing large data sets and encourage them to focus on unbiased and accurate datasets. One could envisage regulatory oversight for certain datasets if the risk of discrimination has its main origin there.

Finally, if algorithms learn by themselves, and develop new solutions to problems, there is a risk that reliance on data results in a pattern of potential discriminatory practices that is learned. If bias is not eliminated or reduced, it will perpetuate the discriminatory decision-making process. Algorithms should strive to be designed in a "gender aware" way, to have less biased outcomes.

3) Concluding Remarks on Datasets and Bias

This paper argues that the focus should be on more accurate and representative data sets instead of falling victim to the *"unreasonable effectiveness of big data"*[132], notably due to its shortcomings. While the quality of the data is key for having accurate, non-discriminatory and fair decisions, this often comes at the cost of inferior algorithms. Establishing a dataset for algorithms[133] is costly for companies which might therefore focus on cheaper and

128 Schauer, Frederick: Profiles, Probabilities, and Stereotypes. Harvard University Press, 2003, p. 131ff.

129 Taleb, Nassim Nicholas: Antifragile: Things that gain from disorder, p. 416–418.

130 Taleb 2012, p. 417.

131 Ibid, p. 417.

132 Halevy, Alon; Norvig, Peter; Pereira, Fernando: The unreasonable effectiveness of data. In: IEEE Intelligent Systems 24.2 (2009), p. 8–12.

133 For example, with the help of humans labeling for example images and using training data for algorithms.

easier methods of collecting massive amounts of data (of poorer quality) for their algorithms. The EC highlighted in its AI White Paper some important elements and requirements for the design of AI systems for (training) data sets, which go in the direction advocated in the present analysis: "Requirements to take reasonable measures aimed at ensuring that such subsequent use of AI systems does not lead to outcomes entailing prohibited discrimination. These requirements could entail obligations to use data sets that are sufficiently representative, especially to ensure that all relevant dimensions of gender [...]are appropriately reflected in those data sets"[134].

At input-level, one solution to overcome biased datasets leading potentially to discrimination by correlation could be a better selection of input data, to "teach" algorithms to avoid bias in the training phase or to let companies and regulators use an algorithm that "checks" the relevant algorithm for biases[135] (a sort of data "TÜV")[136]. Compliance could be either voluntary or mandatory, but it would increase the trust of consumers in algorithms[137].

The challenge is how to detect biases and if a verification should be undertaken for algorithms prior to market entry or only in cases where discrimination occurs. Ferrer et al[138] highlight some of the problems: "To assess whether an algorithm is free from biases, there is a need to analyze the entirety of the algorithmic process. This entails first confirming that the algorithm's underlying assumptions and its modelling are not biased; second, that its training and test data does not include biases and prejudices; and finally, that it is adequate to make decisions for that specific context and task." As discussed earlier, access to information is difficult, notably in the presence of AI. In essence, the source code of the algorithm and the training data is often protected by intellectual property or privacy laws which might prevent training data tests. This complicates identification of biases in the model absent company agreement or a legal provision forcing companies to grant access to the relevant information[139]. Ghili et al. developed a strategy for eliminating (latent) discrimination: "In order to prevent other features

134 European Commission, AI White Paper 2020, p. 19.

135 For algorithms checking datasets and detecting biases, see Bolukbasi et al. 2016.

136 Similar to the German technical inspection association (TÜV) which has the mission to test, inspect and certify technical systems in order to minimize hazards and prevent damages.

137 Increased marketability for companies and possible reputation gains could result.

138 Ferrer et al. 2020, p. 2.

139 Ibid, p. 2.

proxying for sensitive features, we need to include sensitive features in the training phase but exclude them in the test/evaluation phase while controlling for their effects. We evaluate the performance of our algorithm on several real-world datasets and show how fairness for these datasets can be improved with a very small loss in accuracy"[140]. This strategy seems to address the problem of substituting protected characteristics by other proxies while at the same time preserving accuracy. Legislators depend on the development and advancement of such bias detecting techniques by computer scientists to build the corresponding legal tools and adapt the legal framework and enforcement accordingly. The CoE AI Recommendation is clear on this point and should serve as inspiration for regulators: "For the purposes of analyzing the impacts of algorithmic systems [...] on the exercise of rights [...], private sector actors should extend access to relevant individual data and meta-datasets, including access to data that has been classified for deletion, to appropriate parties, notably independent researchers, the media and civil society organizations. This extension of access should take place with full respect to legally protected interests as well as all applicable privacy and data protection rules"[141]. Transforming this guidance into binding law, would ensure the above-mentioned verification process by computer scientists and help victims in discrimination cases to bring evidence in court to prove discrimination claims.

IV. Strategies and Legal Tools to Capture and Overcome 'Algorithmic Discrimination by Correlation'

Eliminating algorithmic biases and achieving GE requires a cross-pollination strategy between regulating algorithms and algorithms assisting the regulator, which can be subdivided into three branches which rely on and influence each other: (1) artificial intelligence assisting the regulator, (2) *non-discrimination by design* and (3) *non-discrimination by law*. *First*, regulators need to be equipped with adequate tools (e.g. specifically designed algorithms for regulators that detect discrimination by correlation), to enforce non-discrimination rules by detecting algorithmic discrimination.

140 Ghili et al. 2019.

141 Recommendation CoE CM/Rec (2020), point 6.1.

Second, the question is whether some forms of discrimination should not (only) be solved by law, but already from the outset in the design of the algorithms by coding in a non-discriminatory way[142]. Such a *non-discrimination by design* approach could help to eliminate some forms of discriminatory behavior (albeit not as stand-alone solution but rather as a complement to traditional regulation), but it raises the question of legal certainty and more generally whether *regulation by code,* is the appropriate form of regulation[143]. This shows interaction between the three strategies and a process of cross-pollination between law, code as well as between the regulator and algorithms.

Third, the cornerstone of each regulatory design aiming to capture discrimination by correlation will be built on *non-discrimination by law*. Choices have to be made between a mix of the above three branches of (legal) strategies, but also between ex-ante, ex-post, general or sector-specific regulation[144], in order to adequately address the issues of *discrimination by correlation* and the underlying root causes (treatment of data and biased datasets). The AI literature has been enriched by many different theoretical reflections[145]. While some authors distinguish code-driven and data-driven regulation while anchoring their regulatory suggestions in the rule of law[146], others call for types of non-discrimination by design, non-discrimination by

142 For a similar idea in relation to code and capital, Pistor, Katharina: The code of capital: How the law creates wealth and inequality. Princeton 2020; Hassan, Samer; De Filippi, Primavera: The Expansion of Algorithmic Governance: From Code is Law to Law is Code. In: Field Actions Science Reports (2017), Special Issue 17.

143 See two examples: Hacker, Philipp: Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. In: Common Market Law Review 55.4 (2018), p. 1143–1185 and Hildebrandt, Mireille: Algorithmic regulation and the rule of law. In: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376.2128 (2018), p. 20170355.

144 On the question if and what type of regulation is appropriate, see Petit 2020, p. 240.

145 See Meneceur, Yannick: L'intelligence artificielle en procès, Bruxelles 2020; Wischmeyer, Thomas: Regulierung intelligenter Systeme. In: Archiv des öffentlichen Rechts 143.1 (2018), p.1–66; Hermstrüwer, Yoan: Artificial Intelligence and Administrative Decisions Under Uncertainty. In: Wischmeyer, Thomas; Rademacher, Timo (Eds.) Regulating Artificial Intelligence. Berlin 2020.; Sunstein, Cass R.: Of artificial intelligence and legal reasoning. University of Chicago Law School. In: Public Law & Legal Theory Working Papers № 18, 2001.

146 Hildebrandt 2018, p. 20170355.

law (either arguing no, minimal, or maximal changes to the current legislative framework) or a combination of both code and law[147].

1) Using Algorithms to Detect Violations of Gender-based Discrimination

While literature and debates tend to focus on the regulation of technology, the question of how regulators could use algorithms to detect algorithmic discrimination[148] or even delegate some form of regulatory power to the stage of coding, is often left aside[149]. There are two separate questions. First, the more fundamental question, whether some forms of discriminatory risks could be captured and avoided, if developers of algorithms (were obliged by law to) use data that significantly reduces the risk of discrimination. This would soften the need for regulatory intervention if empirically less discriminatory harm could be proven. The regulator could couple this with a "marketing authorization style" system known for pharmaceuticals[150] and would merely check the algorithm against incorporated biases before it can be used[151]. Second, whether algorithms should be used by anti-discrim-

147 Cazals, François ; Chantal Cazals: Intelligence artificielle: l'intelligence amplifiée par la technologie. Louvain-la-Neuve 2020, p. 200.

148 Kleinberg et al. 2020.

149 On the question of how algorithms can support regulators, see Alarie, Benjamin; Anthony Niblett; Albert Yoon: Regulation by machine, https://dx.doi.org/10.2139/ssrn.2878950 (February 6, 2021).

150 Detela, Giulia; Lodge, Anthony: EU regulatory pathways for ATMPs: standard, accelerated and adaptive pathways to marketing authorization. In: Molecular Therapy-Methods & Clinical Development 13 (2019), p. 205–232.: "The Marketing Authorisation Application (MAA) procedure [...] ensures the quality, safety, and efficacy of all medicinal products [...] by requiring regulatory review of quality, safety, and efficacy data generated during clinical [...]" which "must comply with the particular standards and requirements within the legislation and the principles of good clinical practice and Good Manufacturing Practice to ensure that the data presented [...] are complete, accurate, and satisfactory."

151 One could model such a system on the market authorization procedure used in pharmaceutical law. Here the clinical trials are also conducted and financed by the industry, underpinned by studies. Companies could test algorithms and datasets with algorithms for biases, submit results to the regulator. That way the industry does the verification and checks itself and the regulator tests and reviews the submitted evidence.

ination bodies to detect discriminatory behavior[152]. "The fact that AI can pick up on discrimination suggests it can be made aware of it. For instance, AI could help spot digital forms of discrimination, and assist in acting upon it."[153] Empowering regulators with algorithmic capabilities would improve and would facilitate decision-making in administration, notably for AI-based discrimination. This might raise questions of knowledge and skills to interpret any such findings, where techniques and algorithms are used to detect biases[154]. For the latter it raises the question to what degree algorithms are merely assisting the regulator with guidance, decision support or substituting (part of) the administrator's decision, in other words, what is the degree of control that the administration would have over the decision-making process. Surely it could not "delegate" regulatory power completely to AI, so control needs to be ensured. Mere assistance to resolve complex matters involving artificial intelligence could be imagined. This is how the concept of cross-pollination is to be understood: algorithms pose challenges and risks for the regulator but also comes along with opportunities, where the power of algorithms can be used to detect discriminatory practices and prove them in court[155]. Regulators could be assisted by algorithms to detect discrimination which could improve the decision-making process. Humans decide differently from machines. While machines are better at abstract and cognitive decision making[156], such as pattern recognition, humans excel at non-abstract decision making such as implicit know-how and intuition as well as ethical and fair reasoning. In this way a combination in the form of support by artificial intelligence while humans remain in the driving seat for the final decision is probably the best mix for taking administrative decisions regarding GE law.

152 See Veale, Michael; Brass, Irina: Administration by algorithm? Public management meets public sector machine learning. In: Public Management Meets Public Sector Machine Learning, Oxford 2019, p. 121–122.; Kleinberg, Jon; Ludwig, Jens; Mullainathan, Sendhil; Sunstein, Cass R.: Discrimination in the Age of Algorithms, In: Journal of Legal Analysis, 10 (2018), p. 113–174.

153 Ferrer et al. 2020, p. 2.

154 Criado Pacheco, Natalia; Ferrer Aran, Xavier; Such, José Mark: A Normative approach to Attest Digital Discrimination. In: Advancing Towards the SDGS Artificial Intelligence for a Fair, Just and Equitable World Workshop of the 24th European Conference on Artificial Intelligence (ECAI'20): AI4EQ ECAI2020.

155 See Hacker 2018.

156 Coeckelbergh 2020, p. 201.

Algorithms often lack accuracy[157] as the run to cost effectiveness often leads to unfair outcomes. If for example, reputation is associated with key words (such as Elite Universities), then representations in the data might have a disproportionate impact on the decision outcome. With speed and pattern recognition algorithms could assist the regulator with sorting and treating cases of discrimination more efficiently. Jointly, algorithms and human decision makers could filter information for investigations, filtering evidence, verify more easily, improve case law analysis, and create more user-friendly data bases. Combining human intelligence and artificial intelligence would probably reduce both type-1 and type-2 errors. A type 1 error (false positive) in our context would be that due to a generalization, a person is being discriminated even though there was no objective reason to discriminate the person. A type 2 error (false negative) would be a situation where a person is not being discriminated (access to credit despite financial problems) despite objective reasons indicating a justification for discriminating a person in terms of access to credit[158].

Relying on generalizations based on the available data entails the risk of decision errors. If generalization is used (due to cost effectiveness), a comparison needs to be made between false positives and false negatives. Such an approach can only be acceptable and justifiable if the errors it produces do not lead to the detriment of the discriminated person. If for example statistical information is available, showing that persons from a specific group with specific characteristics (gender, postal code, attendance of a specific university etc.) typically tend to not repay their loans for example, this could serve as a justification to "label" them with a specific risk and exclude them from receiving for example a credit. However, these practices could lead to exclude persons who despite fulfilling the stereotypical characteristics never had problems paying back a credit and are financially well of.

To conclude, I argue that algorithms assisting the regulator should not be confused with delegating decision-making powers to algorithms. If the regulator remains in control over the decision-making process and artificial intelligence is only assisting human intelligence, there is scope for a better enforcement of gender-based discrimination by correlation, as the potential

157 Chmait Nader; Dowe, David L.; Li Yuan-Fang; Green, David G.: An Information-Theoretic Predictive Model for the Accuracy of AI Agents Adapted from Psychometrics. In: Everitt Tom; Goertzel Ben; Potapov, Alexey (Eds.) Artificial General Intelligence. AGI, Melbourne 2017.

158 For a good example on false negatives and false positives see, Fry 2018, p. 73.

of both systems (artificial and human intelligence) is used to the benefit of fairness and non-discrimination[159]. Therefore, in the spirit of obey-and disobey, not only regulators, but also consumers should be equipped with algorithms (for example put at the disposal by the regulator to enable consumers to test/detect discriminatory behavior) to detect violations of gender-based discrimination which could lead to democratization and decentralization of part of the detection process and lead to better enforcement[160].

2) Non-Discrimination by Design

One could try to reflect as good as possible the principle of non-discrimination at the stage of developing and coding the algorithms. It should never be a substitute to regulation and enforcement but a promising complement to avoid some of the discriminatory behavior. One of the key prerequisites is risk awareness of gender-based discrimination among programmers building AI. Equally to achieving more representative data considering diversity of society, one could address the female gender gap for AI scientists[161] and developers to influence the design and the way algorithms work[162]. A more equal representation of women might shape algorithms for the better. Furthermore, if coders do not understand the legal concept of discrimination, it will be difficult to reflect it in the code. Even if a full "translation" of the concept of discrimination into code will be challenging, basic elements of non-discrimination could be incorporated so that the algorithm tries to avoid or reduce potential discriminations. This could be done for example by checking the datasets used to ensure that they are representative and regularly updated. Some successful processes have been developed in computer science

159 Kleinberg et al. 2020, p. 30097: "The risk that algorithms introduce is not from their use per se, but rather the risk that our regulatory and legal systems will not keep pace with the changing technology."

160 In the absence of a clear legal framework or the lack of algorithms in the regulator's hand (but also if such a system is in place, to support the regulator's enforcement), it could be imagined that consumer rights groups, NGOs, academics and computer scientists could reveal discriminatory algorithms.

161 According to Russel; Norvig 2021 and the "Standford AI100 study" which includes information on diversity, 80% of all AI professors in the world, PhD students and industry hires in the field of AI are male and only 20% female, see Russel; Norvig 2021, p. 45.

162 Crawford 2020, p. 8–9; European Commission, Gender Equality AI Opinion (2020), p. 8–9; Wooldridge 2020, p. 291.

to mitigate risks of bias in datasets, which could tame discriminatory effects already at the stage of programming algorithms.[163]

Finally, despite promising efforts being made to achieve non-discrimination by design, the legal regime continues to apply, and discriminations can be detected and brought to the attention of regulators. The complementary nature of coding the principle of non-discrimination into algorithms is therefore an appreciated effort to help tackle the issue of gender-based discrimination as it also helps regulators. But the need for regulation remains and is even advocated by computer scientists.[164]

3) Non-Discrimination by Law

When choosing to treat the problem of *discrimination by correlation* with the tools of the law, one has several regulatory options. The force of the law, by controlling the behavior of market actors (making them "obey" to legal norms) can influence the AIs' behavior and is certainly the strongest option at the disposal of the state. Regulators often recur to the law because self-regulation and other soft law measures do not solve the problem adequately (a). They have the choice between ex-ante (b) and ex-post (c) regulation. The current analysis has revealed that due to the increasing importance of algorithms a more mutual understanding and exchange on the theoretical side between computer scientist and lawyers on the one hand and between AI programmers/developers as well as business developing an algorithm and regulators is necessary.

a) The Failure of Self-Regulation and Soft Law

In light of not-successful self-regulation and a certain "disobedience" of market actors towards regulators, eliminating biases and stereotypes from datasets is the adequate regulatory approach[165]. Market players lack incen-

163 Celis, L. Elisa; Vijay, Keswani; Vishnoi, Nisheeth: Data preprocessing to mitigate bias: A maximum entropy based approach. In: International Conference on Machine Learning. PMLR, 2020.

164 See for example Mitchell 2019, p. 150–152.

165 For the idea of a regulatory market (albeit for AI safety), see Clark, Jack; Hadfield, Gillian K.: Regulatory Markets for AI Safety (2019). In: arXiv, https://arxiv.org/abs/2001.00078 (February 6, 2021).

tives to ensure that their algorithms don't discriminate as this entails costs for them. In 2019, the OECD adopted a Recommendation on AI, which highlights in its section on human-centered values and fairness, that "AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognized labor rights. "[166]. Despite numerous AI principles published by companies (e.g. Google[167]) and recommendations on AI regulation[168], the OECD principles shall be understood as a call for regulation in order to achieve the objectives laid down in those principles. Importantly, the OECD highlights the observance of the principle of non-discrimination along the whole algorithm lifecycle, which includes industry development, training, data collection, usage, etc. I argue in favor of such an approach which is equally reflected in the CoE Recommendation on AI:" Private sector actors that design, develop or implement algorithmic systems should follow a standard framework for human rights due diligence to avoid fostering or entrenching discrimination throughout all life-cycles of their systems. They should seek to ensure that the design, development and ongoing deployment of their algorithmic systems do not have direct or indirect discriminatory effects on individuals or groups that are affected by these systems, including on those [...] who may face structural inequalities in their access to human rights"[169].

166 OECD, C/MIN (2019)3/FINAL, https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf.

167 Google for example states in point 2 on (unfair) bias, that "AI algorithms and datasets can reflect, reinforce, or reduce unfair biases. recognize that distinguishing fair from unfair biases is not always simple, and differs across cultures and societies. will seek to avoid unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender [...]. ", https://ai.google/principles/ (February 6, 2021).

168 Google has also some suggestions for regulation: "Where existing discrimination laws provide clear guidelines and accountability mechanisms, new rules may be unnecessary. But not all unfair outcomes are the result of illegal discrimination, and some AI systems may have unfair impacts in ways that are not anticipated by existing laws and regulatory frameworks. In these situations, regulators should take a nuanced approach, ensuring that organizations consider the unique historical context in which an AI system is deployed, and use appropriate performance benchmarks for different groups to ensure accountability", https://ai.google/static/documents/recommendations-for-regulating-ai.pdf (February 6, 2021).

169 Recommendation CoE CM/Rec (2020)1 on the human rights impacts of algorithmic systems, point 1.4.

Ultimately, the aim is to ensure fair and non-discriminatory algorithms that improve and facilitate the life of consumers and that earn the companies a fair profit for the innovation and investments. Considering that algorithms usually make "ordinary transactions faster and more efficient"[170], there is a risk of opposition by the industry for fully fledged regulation, as this would entail costs and time. The European Commission's High-Level Expert Group on AI refers to seven key requirements, among which human agency and oversight, transparency, non-discrimination, and fairness as well as accountability are relevant here[171]. Transparency[172] regarding how algorithms take decisions is often considered as a solution to reduce discriminatory AI. Transparency is thought to lead to better decisions and could help overcome "the lack of transparency (opaqueness of AI) makes it difficult to identify and prove possible breaches of laws, including legal provisions that protect fundamental rights, attribute liability and meet the conditions to claim compensation."[173]. However, ensuring transparency is sometimes hard, due to the possibility to include elements of randomness into the algorithm. Including "noise" (e.g. randomness) into the algorithm at the development stage can ensure fairness[174] because it diminishes the impact of each relevant data point. Noise can also be included when consumers provide a lot of data, thereby "diluting" the risk of discrimination. *Explainability* is often based on transparency considerations and thought to lead to transparency[175]. If a certain algorithm can be explained to consumers, this makes the process transparent, and the consumer can take an informed decision.

Achieving transparency and explainability for consumers is sometimes difficult even for the developers. Consumers will not be able to understand the algorithm or why a decision has been taken in a particular way. If pro-

170 Pasquale 2015, p. 213.

171 See European Commission, COM (2019) 168 final, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019DC0168&from=GA.

172 Recommendation CoE CM/Rec (2020)1 on the human rights impacts of algorithmic systems, point 4.1.

173 AI White Paper, p. 14–15: "The opacity of systems based on algorithms could be addressed through transparency requirements".

174 Hosanagar 2020, p. 203.

175 The Coe Recommendation of the CoE on human rights impacts of algorithmic systems include it under transparency in point 4.1. "The use of algorithmic systems in decision-making processes that carry high risks to human rights should be subject to particularly high standards as regards the explainability of processes and outputs."; European Commission, White Paper AI, p. 5.

grammers introduce noise/randomness to ensure fairness it further complicates attempts to ensure both explainability and transparency. In addition, transparency and explainability is often ensured via terms and conditions or user's agreements which must be accepted prior to using a specific service power by AI. These legal constructs are hardly diligently read, except for lawyers or consumer rights groups or in the case of litigation. Consequently, the question is raised whether it is enough and acceptable to delegate a relevant concern of transparency and explainability into the "hidden" corner of terms and conditions.

A similar question of informed consent is also raised in privacy law regarding whether transparency and full information serve the interest of the consumer or whether his or her goal is merely not to be discriminated. Some authors argue for example in the context of privacy regulation, that informed consent in the form of providing information to gain consent is not the appropriate tool to ensure the protection of privacy rights[176]. *In fine*, one could argue that transparency and explainability on their own are not enough and cannot replace regulation. The same holds for AI principles that are only self-binding guidelines for the companies and cannot be enforced in courts.

b) Various Hybrid Models and Ideas of Regulation

There are concrete ideas that blend the approaches of non-discrimination by design/code and non-discrimination by law. One of them has been presented by Hildebrandt as "Ambient Law", "which advocates a framework of technologically embedded legal rules that guarantee transparency of profiles that should allow European citizens to decide which of their data they

176 See for example in that sense, Hermstrüwer, Yoan; Dickert, Stephan: Sharing is daring: An experiment on consent, chilling effects and a salient privacy nudge. International Review of Law and Economics 51 (2017), p. 38–49: "Our study hints at a regulatory dilemma, which arises from the fact that current privacy laws are designed to steer consent choices through salient information and notice: instead of empowering people to make a free and Informed choice over consent, salient information and consent options may push people into conformity. Lawmakers and lawyers might want to consider this risk of backfire effects in the implementation of information and notice policies" and "there is a risk that salient and incentivized consent architectures will systematically push people towards consent with short-term monetary benefits and long-term costs to liberty."

want to hide, when and in which context"[177]. Another approach modelled on computer science wants to incorporate a legal perspective into the design and functioning of algorithms[178]. From a procedural point of view, the detection of discriminatory algorithms could be supported by detection tools made available by the state as open source to support anti-discrimination enforcement, which could serve as complimentary enforcement and due to its decentralized nature would accelerate and facilitate the regulator's effort to detect violations of GE law.

c) Regulating with the Force of the Law: Between Ex-Ante, Ex-Post and Sector Specific Regulation

In light of the high risk of (gender-based) discrimination, there is an argument to regulate algorithms before they enter the market, e.g. via an authorization mechanism (ex-ante). The state could alternatively wait for more information, learn about problems and risks for discrimination of the technology, before intervening (ex-post)[179] ?

There are good arguments on both sides. Rather than regulating in advance and potentially harming and delaying innovations in AI, a more refined approach of regulation could consist in allowing the market forces to do their work but to carefully supervise and regulate as and when market failures or discrimination occurs. Others advocate sector specific regulation[180] rather than general rules, arguing that "Even for algorithmic systems that make decisions about humans, the risks are different in different sectors, and different rules should apply."[181]. Another challenge is the attribution of responsibility for the decisions taken by the algorithm and defining the addressee of the regulatory intervention[182].

177 Hildebrandt, Mireille: Profiling and AmI. In: Rannenberg K., Royer D., Deuker A. (Eds.) The Future of Identity in the Information Society. Berlin, Heidelberg 2009.

178 Criado Pacheco et al. 2020.

179 On the choice between ex-ante and ex-post regulation with a plaidoyer for ex-ante regulation, see Galle, Brian: In Praise of Ex Ante Regulation. In: Vanderbilt Law Review 68 (2015), p. 1715.

180 Zuiderveen 2020, p. 1573.

181 Ibid, p. 1585.

182 See Coeckelbergh, Mark: Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. In: Science Engineering Ethics 26 (2020), p. 2051–2068; Hacker 2018, p. 243–288.

In terms of substantive law, the appropriate design of legal rules needs to capture the behavior causing *discrimination by correlation*. Referring to the work of Searle[183] who claims that a computer program can basically give all outputs desired based on given inputs, which can even be correct, without understanding what it is doing, one could argue that this suggests that computer programs don't possess "intentionality", which only humans can have. This distinguishes computers from humans and is important for the analysis, also with regard to how to assess the discriminatory impact caused by algorithms. In other words, algorithms do not possess "meaning", because meaning is human and can only be given by and expressed by humans. Indirect (discrimination) not requiring the element of knowledge and intent under EU law can be considered an advantage in the context of *discrimination by correlation*.

Regarding the design, the law is always confronted with the dilemma of using generalizations as much as possible to find a rule that captures as many situations as possible instead of regulating many different individual cases (that are unknown in advance). The emergence of a general law of artificial intelligence sees itself confronted with a fast-moving regulatory target[184].

When analyzing potential algorithmic discriminations and statistical data[185] ("statistical discrimination"[186]), generalization plays a more important role than in traditional cases of discrimination.[187] This owes to data mining, huge amounts of data and the classification of consumers into groups to facilitate decision-making. In practice, algorithms discriminate automatically based on (personal) data instead of a specific characteristic of gender[188] and thereby enlarges the field of potential "hooks" to discriminate because many more data points are used compared to the offline world.

183 Searle, John R.: Minds, brains, and programs. In: The Behavioral and Brain Sciences (1980), 3(3), p. 417–424.

184 Barfield, Woodrow; Pagallo, Ugo (Eds.): Research Handbook on the Law of Artificial Intelligence. Cheltenham 2018.

185 For statistical discrimination, association and correlation see Spiegelhalter, p. 109ff.

186 Bohnet, Iris. What works. Boston 2016, p. 31–35, gives an example of statistical discrimination between women and men in negotiations for car sales.

187 Schauer, Frederick: Introduction: The Varieties of Rules. In: Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life. Oxford 1993.

188 See in general, Criado, Natalia; Such, Jose M.: Digital discrimination. Algorithmic Regulation. Oxford 2019.

Any approach could be modelled on existing competition or privacy law[189] enforcement mechanisms[190]. Notably the experience of competition authorities[191] combined with elements from the "newer" approach of data protection enforcement could inform the regulatory approach for *discrimination by correlation*[192].

V. Conclusion

The present analysis has revealed some of the challenges, opportunities and strategies to avoid biased data sets and gender-based discrimination caused by algorithms, which I call *'discrimination by correlation'*.

Several (legal) strategies have been presented that could help overcome or reduce gender bias and discrimination. Even though *non-discrimination by design/non-discrimination by code* that implements the principle of non-discrimination in code is welcomed where feasible, it should be complementary to legislation and regulatory efforts. Strengthening *non-discrimination by law,* it could be envisaged that computer scientists/developers and lawyers/enforcers cooperate on issues of mutual interest and benefit in order to shape the design of algorithms and strive towards the respect of human rights and non-discrimination. An institutionalized forum of exchange of AI and GE experts will enrich both sides and contribute to non-discriminatory AI. Exchange between computer scientists and lawyers should be complemented by including practical knowledge of development and coding also among regulators. Considering that discriminations mostly result from datasets, consumers can also rely on the concept of "noise" and "dis-obey" algorithms by providing a lot of data and introduce elements of randomness.

189 Hacker 2018, p. 5, suggest an interesting approach of combining the enforcement tools of the GDPR-regulation with the concepts of anti-discrimination.

190 For regulatory approach in competition law see, Bailey, Richard; Whish, David: Competition Law, Oxford 2015, p. 1–26.

191 The OECD explores the topic of gender inclusive competition policy by identifying additional relevant features of the market, behavior of consumers and firms, as well as whether a more effective competition policy can help address gender inequality, see http://www.oecd.org/competition/gender-inclusive-competition-policy.htm (February 6, 2021).

192 See specifically on the role of AI and algorithms, Surblytė-Namavičienė, Gintaré: Competition and Regulation in the Data Economy, London 2020.

The law remains the *conditio sine qua non* to ensure the fair, ethical and non-discriminatory use of algorithms. Even though EU non-discrimination law is flexible in principle to deal with some of the challenges[193] arising with *discrimination by correlation*[194], the law needs to evolve in light of technological developments to adequately capture gender-based *discrimination by correlation* and ensure sufficient legal protection to victims of gender-based discrimination.

Literature

Adam, Alison: Artificial knowing: gender and the thinking machine. London 1998.

Agrawal, Ajay; Joshua Gans; Avi Goldfarb: Máquinas predictivas: la sencilla economía de la inteligencia artificial, Madrid 2019.

Alarie, Benjamin; Anthony Niblett; Albert Yoon: Regulation by machine, http://dx.doi.org/10.2139/ssrn.2878950 (February 6, 2021).

Ali, Muhammad; Sapiezynski, Piotr; Bogen, Miranda; Korolova, Aleksandra; Mislove, Alan; Rieke, Aaron (2019): Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes. In: arXiv e-prints, https://arxiv.org/pdf/1904.02095.pdf (February 6, 2021).

Anrig, Bernhard; Browne, Will; Gasson, Mark: The Role of Algorithms in Profiling. In: Hildebrandt, Mireille; Gutwirth, Serge (Eds.): Profiling the European Citizen. Berlin 2008.

Badre, David: On Task, Princeton 2020.

Bailey, Richard; Whish, David: Competition Law, Oxford 2015.

Barfield, Woodrow; Pagallo, Ugo (Eds.): Research Handbook on the Law of Artificial Intelligence. Cheltenham 2018.

193 Agreeing in principle while also suggesting additional regulation to capture algorithmic decision-making, Zuiderveen 2020, p. 1585.

194 See discussion of the CJEU case Schuch-Ghannadan (Section II.) which suggests that in cases of discrimination by correlation, there are good arguments for claimants, who succeed in bringing prima facie evidence for an alleged discrimination in cases where access to evidence is impossible or only unreasonably difficult (which is typically the problem surrounding the workings of the algorithm), that the burden of proof shifts to the company who would need to proof that their algorithm did not discriminate.

Barocas, Solon; Selbst, Andrew D.: Big Data's Disparate Impact. In: California Law Review 104 (2016), p. 671–732.

Barocas, Solon: Data & Civil Rights: Technology Primer (2014), http://www.datacivilrights.org/pubs/2014-1030/Technology.pdf (February 6, 2021).

Basdevant, Adrien; Mignard, Jean-Pierre: L'Empire des données. Essai sur la société, les algorithmes et la loi. Paris 2018.

Boden, Margaret A: AI: Its nature and future. Oxford 2016.

Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James; Saligrama, Venkatesh, Kalaim Adam: Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In: arXiv preprint arXiv:1607.06520 (2016), https://arxiv.org/abs/1607.06520v1 (February 6, 2021).

Bohnet, Iris: What works. Boston 2016.

Bostrom, Nick: Superintelligence. Paris 2017.

Bostrom, Nick: The future of humanity. In: Geopolitics, History, and International Relations 1(2) (2009), 41–78.

Buolamwini, Joy; Timnit Gebru: Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Conference on fairness, accountability and transparency. PMLR, 2018.

Cazals, François ; Chantal Cazals: Intelligence artificielle: l'intelligence amplifiée par la technologie. Louvain-la-Neuve 2020.

Celis, L. Elisa ; Mehrotra, Anay; Vishnoi, Nisheeth: Toward controlling discrimination in online ad auctions. In: International Conference on Machine Learning. PMLR, 2019, https://arxiv.org/abs/1901.10450.

Celis, L. Elisa; Kapoor, Sayash; Salehi, Farnood; Vishnoi K. Nisheeth: An algorithmic framework to control bias in bandit-based personalization. In: arXiv preprint arXiv:1802.08674(2018).

Celis, L. Elisa; Vijay, Keswani; Vishnoi, Nisheeth: Data preprocessing to mitigate bias: A maximum entropy based approach. In: International Conference on Machine Learning. PMLR, 2020.

Chmait Nader; Dowe, David L.; Li Yuan-Fang; Green, David G.: An Information-Theoretic Predictive Model for the Accuracy of AI Agents Adapted from Psychometrics. In: Everitt Tom; Goertzel Ben; Potapov, Alexey (Eds.) Artificial General Intelligence. AGI, Melbourne 2017.

Clark, Jack; Hadfield, Gillian K.: Regulatory Markets for AI Safety. In: arXiv preprint arXiv:2001.00078 (2019), https://arxiv.org/abs/2001.00078 (February 6, 2021).

Coeckelbergh, Mark: AI Ethics, Boston 2020.

Coeckelbergh, Mark: Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. In: Science Engineering Ethics 26 (2020), p. 2051–2068.

Craig, Paul; Gráinne De Búrca: EU law: text, cases, and materials. Oxford 2020.

Crawford, Kate; Whittaker, Meredith; Elish, Madelein Clare; Barocas, Solon; Plasek, Aaron; Ferryman, Kadija: The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term, New York 2016.

Criado Pacheco, Natalia; Ferrer Aran, Xavier; Such, José Mark Coté: A Normative approach to Attest Digital Discrimination. In: Advancing Towards the SDGS Artificial Intelligence for a Fair, Just and Equitable World Workshop of the 24th European Conference on Artificial Intelligence (ECAI'20): AI4EQ ECAI2020.

Criado, Natalia; Such, José Mark: Digital discrimination. Algorithmic Regulation, Oxford 2019.

Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, 8th April 2020, https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154 (February 6, 2021).

DeMichele, Matthew; Baumgartner, Peter; Wenger, Michael; Barrick, Kelle; Comfort, Megan; Misra, Shilpi: The Public Safety Assessment: A re-validation and assessment of predictive utility and differential prediction by race and gender in Kentucky (2018), https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/announcements/predictiveutilitystudy.pdf (February 6, 2021).

Detela, Giulia; Lodge, Anthony: EU regulatory pathways for ATMPs: standard, accelerated and adaptive pathways to marketing authorization. In: Molecular Therapy-Methods & Clinical Development 13 (2019), p. 205–232.

Ellis, Evelyn; Watson, Philippa: EU anti-discrimination law, Oxford 2012.

Ernst, Ekkehardt; Merola, Rossana; Samaan, Daniel: Economics of artificial intelligence: Implications for the future of work. In: IZA Journal of Labor Policy 9.1 (2019), p. 16.

European Commission, Advisory Committee on Equal Opportunities for Women and Men, Opinion on Artificial Intelligence (2020), https://ec.europa.eu/info/sites/info/files/aid_development_cooperation_fundamental_rights/opinion_artificial_intelligence_gender_equality_2020_en.pdf (February 6, 2021).

European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intellgience (Artificial Intelligence Act) and amending certain Union legislative acts, COM (2021) 206 final.

European Commission, "Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law" (2021) prepared by the Legal Network of Gender Experts of the EC, https://www.equalitylaw.eu/downloads/5361-algorithmic-discrimination-in-europe-pdf-1-975

European Commission, White Paper, On Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (February 6, 2021).

European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM/2020/825 final.

European Commission, Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act), COM/2020/842 final.

Ferrer, Xavier; van Neunen, Tom; Such, Jose M.; Coté, Mark; Criado, Natalia: Bias and Discrimination in AI: a cross-disciplinary perspective. In: arXiv preprint arXiv:2008.07309 (2020), p. 1., https://arxiv.org/abs/2008.07309 (February 6, 2021).

Surblytė-Namavičienė, Gintaré: Competition and Regulation in the Data Economy, London 2020.

Fry, Hannah: Hello World: How to be Human in the Age of the Machine, London 2018.

Galle, Brian: In Praise of Ex Ante Regulation. In: Vanderbilt Law Review 68 (2015), p. 1715.

Geerts, Thierry: Homo Digitalis, Tielt 2021.

Ghili, Soheil; Ehsan Kazemi; Amin Karbasi: Eliminating latent discrimination: Train then mask. In: Proceedings of the AAAI Conference on Artificial Intelligence (2019), 33. Nº 01.

Gunkel, David. J.: The Machine Question: Critical Perspectives on AI, Robots, and Ethics, Cambridge 2012.

Hacker, Philipp: Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. In: Common Market Law Review 55.4 (2018), p. 1143–1185.

Hacker, Philipp: Verhaltens- und Wissenszurechnung beim Einsatz von Künstlicher Intelligenz. In: RW Rechtswissenschaft 2018, p. 243–288.

Halevy, Alon; Norvig, Peter; Pereira, Fernando: The unreasonable effectiveness of data. In: IEEE Intelligent Systems 24.2 (2009), p. 8–12, https://dl.acm.org/doi/10.1109/MIS.2009.36 (February 6, 2021).

Hardt, Moritz: How big data is unfair. In: Medium 2014, https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de (February 6, 2021).

Hart, Herbert Lionel Adolphus; Green, Leslie: The concept of law. Oxford 2015.

Hassan, Samer; De Filippi, Primavera: The Expansion of Algorithmic Governance: From Code is Law to Law is Code. In: Field Actions Science Reports (2017), Special Issue 17, http://journals.openedition.org/factsreports/4518 (February 6, 2021).

Hermstrüwer, Yoan; Dickert, Stephan: Sharing is daring: An experiment on consent, chilling effects and a salient privacy nudge. In: International Review of Law and Economics 51 (2017), p. 38–49.

Hermstrüwer, Yoan: Artificial Intelligence and Administrative Decisions Under Uncertainty. In: Wischmeyer, Thomas; Rademacher, Timo (Eds.): Regulating Artificial Intelligence. Berlin 2020.

Hildebrandt, Mireille: Profiling and AmI. In: Rannenberg K., Royer D., Deuker A. (Eds.): The Future of Identity in the Information Society. Berlin, Heidelberg 2009.

Hildebrandt, Mireille: Profiling: From data to knowledge. In: Datenschutz und Datensicherheit DuD (2006) 30, p. 548–552.

Hildebrandt, Mireille: Algorithmic regulation and the rule of law. In: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376.2128 (2018): 20170355.

Hosanagar, Kartik: A human's guide to machine intelligence: how algorithms are shaping our lives and how we can stay in control, New York 2020.

Keller, John D: Deep Learning, Boston 2019.

Kleinberg, Jon; Ludwig, Jens; Mullainathan, Sendhil; Sunstein, Cass R.: Algorithms as discrimination detectors, In: Proceedings of the National Academy of Sciences Dec 2020, 117 (48), p. 30096–30100.

Kleinberg, Jon; Ludwig, Jens; Mullainathan, Sendhil; Sunstein, Cass R.: Discrimination in the Age of Algorithms, In: Journal of Legal Analysis, 10 (2018), p. 113–174.

Kim, Pauline T.: Data-driven discrimination at work. In: William & Mary Law Review 58 (2016), p. 857–866.

Kraft-Buchman, Caitlin; Arian, Réné: The Deadly Data Gap: Gender and Data. Women at the Table, Geneva 2019, http://bit.ly/DeadlyDataGenderGap (February 6, 2021).

Lambrecht, Anja; Tucker, Catherine E.: Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads. In: Management Science, 65 (2019), p. 2966–2981.

Lee, Felicia R.: Discriminating? Yes. Discriminatory? Nº The New York Times, December 13,2003. advance.lexis.com/api/document?collection=news&id=urn:contentItem:4B70-S120-01KN-20KW-00000-00&context=1516831 (February 6, 2021).

Liao, S. Matthew: A Short Introduction to the Ethics of Artificial Intelligence. Ethics of Artificial Intelligence: Oxford University Press, 2020.

Louridas, Panos: Algorithms, Boston 2020.

Mayson, Sandra G.: Bias in, bias out. In: Yale Law Journal 128 (2018), p. 2218.

McAfee, Andrew Paul; Brynjolfsson, Erik: Machine, Platform, Crowd: Harnessing Our Digital Future, New York 2017, p. 67.

Meneceur, Yannick: L'intelligence artificielle en procès, Bruxelles 2020.

Mitchell, Melanie: Artificial intelligence: A guide for thinking humans. London 2019.

OECD, Principles on Artificial Intelligence, https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm (February 6, 2021).

Orwat, Carsten: Risks of Discrimination through the Use of Algorithms. A study compiled with a grant from the Federal Anti-Discrimination Agency. Berlin 2020.

Packin, Nizan; Lev-Aretz, Yafit: Learning algorithms and discrimination. In: Research Handbook on the Law of Artificial Intelligence. Cheltenham, London 2018, p. 91.

Pasquale, Frank: The Black Box Society: The Hidden Algorithms Behind Money and Information, Boston 2015.

Perez, Caroline Criado: Invisible women: Exposing data bias in a world designed for men. London 2019.
Petit, Nicolas: Big Tech and the Digital Economy: The Moligopoly Scenario. Oxford 2020.
Pistor, Katharina: The code of capital: How the law creates wealth and inequality. Princeton 2020.
Renan Barzilay, Arianne: The Technologies of Discrimination: How Platforms Cultivate Gender Inequality. In: The Law & Ethics of Human Rights, 13 (2019), no. 2, p. 179–202.
Russell, Stuart: Artificial intelligence: The future is superintelligent. In: Nature 548 (2017), p. 520–521.
Russell, Stuart; Norvig, Peter: Artificial intelligence: a modern approach, London 2021.
Schauer, Frederick: Introduction: The Varieties of Rules. In: Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life. Oxford 1993.
Schauer, Frederick: Profiles, Probabilities, and Stereotypes. Boston 2003.
Searle, John R.: Minds, brains, and programs. In: The Behavioral and Brain Sciences, 3(3) (1980), p. 417–424.
Skeem, Jennifer; John Monahan; Christopher Lowenkamp: Gender, risk assessment, and sanctioning: The cost of treating women like men. In: Law and human behavior 40.5 (2016), p. 580.
Spiegelhalter, David: The art of statistics: learning from data, London 2019.
Strauß, Stefan: From Big Data to Deep Learning: A Leap Towards Strong AI or 'Intelligentia Obscura'? In: Big Data Cogn. Comput. 2 (2018), p. 16.
Sunstein, Cass R.: Of artificial intelligence and legal reasoning. University of Chicago Law School, In: Public Law & Legal Theory Working Papers № 18, 2001, http://chicagounbound.uchicago.edu/ public_law_and_legal_theory/207/ (February 6, 2021).
Taleb, Nassim Nicholas: Antifragile: Things that gain from disorder, p. 416–418.
Tegmark, Max: Life 3.0: Being human in the age of artificial intelligence, London 2017.
US White House, Executive Office of the President, Big Data: Seizing Opportunities, preserving, Washington 2014, http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report _5.1.14_final_print.pdf (February 6, 2021).
UNESCO, Report on AI and Gender Equality, https://unesdoc.unesco.org/ark:/48223/pf0000374174 (February 6, 2021).

Veale, Michael; Brass, Irina: Administration by algorithm? Public management meets public sector machine learning. Public Management Meets Public Sector Machine Learning, Oxford 2019, p. 121–122.

Wellner, Galit; Rothman, Tiran: Feminist AI: Can We Expect Our AI Systems to Become Feminist? In: Philosophy & Technology 33 (2020), p. 191–205.

Wischmeyer, Thomas: Regulierung intelligenter Systeme. In: Archiv des öffentlichen Rechts 143.1 (2018), p. 1–66.

Wooldridge, Michael: The Road to Conscious Machines: The Story of AI. London 2020.

Wright, Emily M.; Salisbury, Emily J.; Van Voorhis, Patricia: Predicting the prison misconducts of women offenders: The importance of gender-responsive needs. In: Journal of Contemporary Criminal Justice 23.4 (2007), p. 310–340.

Xenidis, Raphaële; Senden, Linda: EU Non Discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination, in: Bernitz, Ulf; Groussot, Xavier; de Vries, Sybe A. (Eds.): General Principles of EU law and the EU Digital Order, Bruxelles 2020.

Zuboff, Shoshana: The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power, New York 2019.

Zuiderveen Borgesius, Frederik: Discrimination, artificial intelligence, and algorithmic decision-making. Study for the Council of Europe, Strasbourg 2018, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 (February 6, 2021).

Zuiderveen Borgesius, Frederik: Strengthening legal protection against discrimination by algorithms and AI. In: The International Journal of Human Rights (2020), 24:10, p. 1572–1593.