CHAPTER 5

Shared Responsibility for Human Rights in the Algorithmic Age: Why Business Should Be the States' Ally to Eliminate Discrimination

Fabian Lütz

1 Introduction

As algorithms transform online and offline worlds for equality, this chapter explores the role of businesses in ensuring a human rights compliant approach to regulating algorithms.[1] States and businesses around the world increasingly invest in Artificial Intelligence ("AI") and adopt and rely on AI systems[2] to make decisions.[3] Biases and discrimination could occur when AI and algorithms are used to complement or substitute human decision-making. For example, biases can occur in credit lending[4] or algorithmic recruitment tools which use datasets to train, validate, test and run the algorithm. Those datasets use historic data, which might contain stereotypes and biases from the real world and therefore lead to biased or discriminatory decisions.[5] Biases can

1 The author would like to thank the organizers and participants of the conference Business and Human Rights in Lausanne (June 30 to July 1, 2023) and in particular Dr. Joseph Wilson for his very useful oral and written comments on the draft chapter. For the sake of simplicity, algorithms and AI systems are used interchangeably. AI systems are understood here as a "machine-based system that for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments" (Art. 2 Council of Europe, [CETS 225] – Artificial Intelligence; Art. 3(1) AI Act).

2 See "The OECD Framework for the Classification of AI systems," Organisation for Economic Co-operation and Development ("OECD"), accessed May 25, 2024, https://wp.oecd.ai/app/uploads/2022/02/Classification-2-pager-1.pdf.

3 See "Artificial Intelligence: ChatGPT Inc," The Economist, Business, July 1, 2023, 48, which is developing AI adoption index by business.

4 See Web Arnold, "ANALYSIS: What Lenders Should Know About AI and Algorithmic Bias," Bloomberg Law, April 25, 2023, https://news.bloomberglaw.com/bloomberg-law-analysis/analysis-what-lenders-should-know-about-ai-and-algorithmic-bias.

5 See Matthew Burgess, *Artificial Intelligence* (*WIRED guides*)*: How Machine Learning Will Shape the Next Decade* (Random House Business, 2021), 77–80; probably the most cited example is the algorithm developed by Amazon. See Jeffrey Dastin, "Insight – Amazon scraps secret AI recruiting tool that showed bias against women," Reuters, October 11, 2018,

 | DOI:10.1163/9789004715158_007

also occur in word associations or embeddings,[6] which are often used for training algorithms or within machine learning algorithms, such as search engines that identify patterns in order to rank or suggest the next word in the autocomplete function of search engines. Another example is the biased distribution of online job advertisements in the framework of micro-targeting which can be biased or discriminatory, like on the basis of race or gender. Finally, similar to the retrieval of existing data with search engines that might be biased or stereotyped, since the increased use of Large Language Models ("LLMs"), biases can also occur in text or images generated by LLMs.[7] In order to reduce such biases and algorithmic discrimination, this chapter argues for businesses and states to have a a shared responsibility based on human rights frameworks and gives some elements that should form part of any legislative framework. Considering the global nature of AI, multinational enterprises[8] and international organizations play a primordial role in shaping the human rights framework on AI. The G7 recently called for responsible AI and global governance.[9]

The Toronto Declaration, which aims to protect the right to equality and non-discrimination in machine learning systems, clearly sets out to hold private sector actors to account by specifically stating "[i]nternational law clearly sets out the duty of states to protect human rights; this includes ensuring the right to non-discrimination by private sector actors."[10] Calls for regulating AI

---

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

6 Tolga Bolukbasi, et al., "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings," arXiv, July 21,2016, https://arxiv.org/abs/1607.06520.

7 See for example the generation of images of CEO with the prompt "CEO" which creates images of mostly men with light skin color; see Leonardo Nicoletti and Dina Bass, "Humans are Biased. Generative AI is Even Worse," Bloomberg, June 9, 2023, https://www.bloomberg.com/graphics/2023-generative-ai-bias/?srnd=equality#xj4y7vzkg, notably on Stable Diffusion's text-to-image model which produced biased and stereotyped images in relation to gender and race.

8 See for example the "OECD Guidelines for Multinational Enterprises on Responsible Business Conduct," OECD, June 8, 2023, paras. 5, 51, and 54, https://www.oecd-ilibrary.org/finance-and-investment/oecd-guidelines-for-multinational-enterprises-on-responsible-business-conduct_81f92357-en and "Recommendation of the Council on Artificial Intelligence," OECD, amended May 3, 2024, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

9 "Ministerial Declaration: The G7 Digital and Tech Ministers' Meeting – 30 April 2023," G7/G20, April 30, 2023, paras. 39–48, https://g7g20-documents.org/database/document/2023-g7-japan-ministerial-meetings-ict-ministers-ministers-language-ministerial-declaration-the-g7-digital-and-tech-ministers-meeting.

10 "The Toronto Declaration: Protecting the right to equality in machine learning," Toronto Declaration, May 16, 2018, para. 38, https://www.torontodeclaration.org/declaration-text/english/ ("Toronto Declaration").

and holding AI companies that create algorithms accountable are increasingly heard, which is not surprising considering companies make increasing use of AI systems such as ChatGPT or GPT-4[11] and different scholars have raised concerns over the dangers of an *algocracy*.[12] The UN Committee on Economic, Social and Cultural Rights has equally emphasized that "States parties must [...] adopt measures, which should include legislation, to ensure that individuals and entities in the private sphere do not discriminate on prohibited grounds."[13] Within this spirit, the Toronto Declaration insists that

> States should put in place regulation compliant with human rights law for oversight of the use of machine learning by the private sector in contexts that present risk of discriminatory or other rights-harming outcomes, recognizing technical standards may be complementary to regulation. In addition, non-discrimination, data protection, privacy and other areas of law at national and regional levels may expand upon and reinforce international human rights obligations applicable to machine learning.[14]

Relying on UN reports and policy proposals on race, disability, and gender as well as existing and proposed human rights frameworks on AI, this chapter discusses biased decision-making and discriminatory outcomes of AI. Rather

---

11 Jules Thomas et al., "De ChatGPT à Midjourney, les intelligences artificielles génératives s'installent dans les entreprises," LeMonde April 26, 2023 https://www.lemonde.fr/economie/article/2023/04/25/de-chatgpt-a-midjourney-les-intelligences-artificielles-generatives-s-installent-dans-les-entreprises_6170873_3234.html.

12 See for example, Arthur Grimonpont, *Algocratie: Vivre libre à l'heure des algorithmes* (Actes Sud, 2022); Hugues Bersini and Gilles Badinet, *Algocratie: Allons-nous donner le pouvoir aux algorithmes?* (De Boeck Supérieur, 2023).

13 "General Comment No. 20, Non-discrimination in economic, social and cultural rights (arts 2, para. 2, of the International Covenant on Economic, Social and Cultural Rights)," UN Committee on Economic, Social and Cultural Rights ("CESCR"), July 2, 2009, para. 11, https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-20-2009-non-discrimination.

14 Toronto Declaration, *supra* note 10, at para. 40. See in this context also Art. VIII of the Spanish Carta Derechos Digitales: "Derecho a la igualdad y a la no discriminación en el entorno digital: 1. El derecho y el principio a la igualdad inherente a las personas será aplicable en los entornos digitales, incluyendo la no discriminación y la no exclusión. En particular, se promoverá la igualdad efectiva de mujeres y hombres en entornos digitales. Se fomentará que los procesos de transfor- mación digital apliquen la perspectiva de género adoptando, en su caso, medidas específicas para garantizar la ausencia de sesgos de género en los datos y algoritmos usados." "Carta Derechos Digitales," La Moncloa, July 24, 2021, Article VIII, https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf.

than being neutral or value-free, algorithms are only as "good" as the underlying data and the design choices of their creators. Both data and design can reflect existing inequalities and stereotypes of society, reinforce discrimination, or facilitate multiple or intersectional discrimination. Algorithmic discrimination is understood here as an unjustified difference in treatment on the basis of a protected characteristic (such as gender or race), where a human decision is either fully or partially replaced or substantially assisted by an algorithm which causes a discrimination for a human.[15]

First, assuming regulating AI is a *conditio sine qua non*, the author sketches out the role of business in supporting and ensuring a human rights compliant approach to avoid discrimination (Section 3). Self-binding AI principles and standards of companies are mushrooming worldwide (Section 2). In parallel, international organizations and governments are proposing regulation on AI to protect human rights (Section 4). The analysis will be based on a review and discussion of selected relevant self-binding AI standards of business as well as soft law and legal proposals that foresee a specific role of businesses to achieve human rights in the algorithmic context.[16]

15 See also the proposed definition of algorithmic discrimination in the US Blueprint for an AI Bill of Rights: "'Algorithmic discrimination' occurs when automated systems contribute to unjustified different treatment or impacts disfavoring people based on their race, color, ethnicity, sex, religion, age, national origin, disability, (..) or any other classification protected by law." "Blueprint for an AI Bill of Rights," The White House, last modified November 22, 2023, https://www.whitehouse.gov/ostp/ai-bill-of-rights/definitions/.

16 See for example "Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems," Council of Europe ("CoE"), April 8, 2020, https://rm.coe.int/0900001680 9e1154; "Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie – Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework," UN Human Rights Council ("HRC"), March 21, 2011, https://digitallibrary.un.org/record/705860?ln=en&v=pdf ("UNGPs"); "Rights of persons with disabilities: Report of the Special Rapporteur on the rights of persons with disabilities," HRC, December 28, 2021, https://www.ohchr.org/en/documents/thematic-reports/ahrc4952-artificial-intelligence-and-rights-persons-disabilities-report; "Racial discrimination and emerging digital technologies: a human rights analysis – Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*," HRC, June 18, 2020, https://www.undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2F44%2F57&Language=E&DeviceType=Desktop&LangRequested=False; "The right to privacy in the digital age*, Report of the United Nations High Commissioner for Human Rights," HRC, September 13, 2021, https://www.ohchr.org/en/documents/thematic-reports/ahrc4831-right-privacy-digital-age-report-united-nations-high; "Digital Innovation, technologies and the right to health: Report of the Special Rapporteur on the right of everyone to the enjoyment of hte highest attainable standard

Second, against the background of a risk of an (in)visible shift from classical public lawmaking towards private rule-setting, this chapter critically assesses the advantages and limits of including businesses in private regulatory tasks to avoid discrimination. Arguing for a (human) rights-based approach as fundament for regulation, it acknowledges the essential role of business in reducing biases and discrimination and recognizes that in specific situations, such as very competitive markets with a demand by consumers for non-discriminatory AI, less regulatory intervention may be required.

In conclusion, the chapter argues for recognizing a shared responsibility of businesses and states for human rights in the algorithmic age. It identifies and sketches out their respective roles and regulatory approaches to achieve less biases and discrimination (see Section 6 on elements and recommendations for a potential "shared responsibility" framework). Involving businesses as partners and addressees of obligations imposed by legal frameworks is more important than ever in a world where AI technologies, such as LLMs,[17] are developed and deployed more quickly than regulators are able to adopt or adapt new binding rules.[18] In any case, while waiting for new general or specific AI rules to be adopted, regulators can rely on general principles of law or specific laws dealing with the technological issues, to investigate and deal with algorithmic discrimination to the extent possible.[19] UN experts recently recalled the urgent need for regulation and transparency with regard

---

of physical and mental health," HRC, April 23, 2023, https://www.ohchr.org/en/documents/thematic-reports/ahrc5365-digital-innovation-technologies-and-right-health, which recalls for example that "the same rights that are protected offline must be protected with the use of digital tools and in the online world."

17 "An LLM is a computerized language model, embodied by an artificial neural network using an enormous amount of 'parameters' that are (pre-)trained on many GPUs in relatively short time due to massive parallel processing of vast amounts of unlabeled texts (..);" see "Large language model," Wikipedia, accessed May 25, 2024, https://en.wikipedia.org/wiki/Large_language_model; See also David Nield, "How ChatGPT and Other LLMs Work – and Where They Could Go Next," WIRED, April 30, 2023, https://www.wired.com/story/how-chatgpt-works-large-language-model/.

18 Therefore, regulators need to adopt sufficiently broad definitions and principles that capture potential future AI developments and that at the same time rely on specific and detailed requirements for regulating AI systems. The EU, for example, relies on a regulatory technique that enables the European Commission to change the Annex to the Regulation and thereby ensures a dynamic regulatory system that is fit for purpose to adapt flexibly to new arising AI systems.

19 See for example the policy recommendation in the report "Ghost in the Machine," Norwegian Consumer Council, June 2023, 60, https://storage02.forbrukerradet.no/media/2023/06/generative-ai-rapport-2023.pdf: "[e]nforcement agencies must not wait for upcoming regulation. Instead, they must immediately investigate generative AI systems

to new and emerging technologies highlighting the non-discrimination and gender angle.[20] However, shared responsibility should not lead to regulatory capture. Relying on regulation always entails the risk of regulatory capture,[21] especially in areas where technical knowledge is essential to understanding the regulatory object. As a recent example from the aviation industry has shown.[22]

2 Self-binding AI Principles and Ethical Standards (Soft Law)

Without much exaggeration, one can speak of the mushrooming of AI principles, best practices, and ethical guidelines on AI.[23] The reasons for this phenomenon, whether they are window-dressing, ethics washing, pure marketing, or conscious and deliberate efforts to address the issues at the company level, will not be discussed. What is of interest in the following analysis is the topics addressed by those non-binding guidelines (2.1) and the limits and shortcomings of mere soft law instruments when it comes to addressing algorithmic discrimination (2.2). Even though non-binding in nature, some principles or

---

and apply relevant legal provisions from their respective legal frameworks, such as data protection, competition, product safety and consumer law."

20 See "New and emerging technologies need urgent oversight and robust transparency: UN experts," Office of the United Nations High Commissioner for Human Rights ("OHCHR"), June 2, 2023, https://www.ohchr.org/en/press-releases/2023/06/new-and-emerging-technologies-need-urgent-oversight-and-robust-transparency.

21 See for example Richard A. Posner, "The Concept of Regulatory Capture: A Short, Inglorious History," in *Preventing Regulatory Capture: Special Interest Influence and How to Limit it*, eds. Daniel Carpenter and David A. Moss (Cambridge University Press, 2013), 49–56.

22 A recent example of potential regulatory capture is the US Federal Aviation Administration's reliance on Boeing's engineers in certifying Boeing 737 MAX planes, where the faulty Maneuvering Characteristics Augmentation System (MCAS) which provides consistent airplane handling characteristics in a very specific set of unusual flight conditions resulted in two crashes. See for example the report of the Ethiopian Ministry investigating the causes of the plane crash. "Investigation Report on Accident to the B737-MAX8 Reg. ET-AVJ Operated by Ethiopian Airlines, 10 March, 2019," The Federal Democratic Republic of Ethiopia Ministry of Transport and Logistics Aircraft Accident Investigation Bureau, December 23, 2022, https://bea.aero/fileadmin/user_upload/ET_302__B737-8MAX_ACCIDENT_FINAL_REPORT.pdf.

23 Anna Jobin, Marcello Ienca, and Effy Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence* 1, no. 9 (September 2019): 389–99. See also Helga Nowotny, *In AI We Trust: Power, Illusion and Control of Predictive Algorithms* (Polity Press, 2021), 123–25.

social impact statements that include guidance for avoiding biases and non-discrimination could actually bring about meaningful change if relied upon by developers in their daily coding activities as guidance (3.).[24]

2.1 *The Main Principles in Non-binding Guidelines*

Companies follow AI principles that typically address some of the main topics of equality and non-discrimination using wording derived from the diversity, inclusion, equality, and non-discrimination context.[25] As such, most guidelines issued by the major AI companies (GAFAM)[26] address human rights issues[27] but tend to avoid specific issues of gender equality or non-discrimination.

As a matter of illustration, *diversity and inclusion*, which is better perceived as a political as opposed to a legal objective, needs to be addressed by policies rather than in a legal framework. One company stated that despite general consensus, reports and studies by public bodies that demonstrate a lack of diversity among coders and developers[28] in AI companies that:

---

24 For example, Nicholas Diakopoulos et al., "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms," FAT/ML, accessed May 25, 2024, https://www.fatml.org/resources/principles-for-accountable-algorithms.

25 See for example Annie Batlle, Aude Bernheim, and Flora Vincent, *L'intelligence artificielle, pas sans elles!* (Belin, 2019).

26 "Our Principles," Google, accessed May 25, 2024, https://ai.google/principles/; "Transformieren Sie verantwortungsvolle KI von der Theorie in die Praxis," Amazon, accessed May 25, 2024, https://aws.amazon.com/de/machine-learning/responsible-machine-learning/; "Facebook's five pillars of Responsible AI," Meta, June 22, 2021, https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/; "Microsoft Responsible AI Standard, v2," Microsoft, June 2022, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl. Under the DSA, some of the services of Google (Play, Maps, Shopping), Apple (AppStore), and Amazon (Store) have been designated as "Very Large Online Platforms" which result in additional compliance requirements. Equally, Bing (Microsoft) and Google Search have been designated as Very Large Online Search Engines. See "Digital Services Act: Commission designates first set of Very Large Online Platforms and Search Engines," European Commission, April 25, 2023, https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413.

27 Some even issue general human rights reports; see, for example, Miranda Sissons, "A Closer Look: Meta's First Annual Human Rights Report," Meta, July 14, 2022, https://about.fb.com/news/2022/07/first-annual-human-rights-report/.

28 Kimberly A. Houser, "Can AI solve the Diversity Problem in the Tech Industry: Mitigating Noise and Bias in Employment Decision-Making," *Stanford Technology Law Review* 22 (February 2019): 290–354; Susan Leavy, "Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning," paper presented at the Proceedings of the 1st international workshop on gender equality in software engineering, May 27 – June 3, 2018, https://ieeexplore.ieee.org/document/8452744.

> We are also prioritizing AI Diversity & Inclusion education efforts for our AI team when hiring and training employees, and setting clear D&I expectations for our AI managers. We aim to better ensure that the people making our AI products are from as diverse a range of backgrounds and perspectives as the people using them, and that we are inclusive of a broad range of voices in our decision-making.[29]

It is good that companies address these issues in their AI principles, but in practice they also need to follow up on those promises to ensure that diverse mindsets of developers lead to less biased algorithm designs. Such general statements in AI company policies resemble marketing statements and should similarly be complemented by measurable targets, such as achieving 40% representation of underrepresented groups. Statistics on the representation of women in AI companies could inform customers of these AI projects whether the priority goals are pursued in hiring and training. In addition, the decision-making process should be made more transparent in order to verify whether the AI lifecycle indeed includes the broad range of backgrounds claimed in the company policy.

On *biases* Google outlines its position on avoiding biases in its AI Principles under the heading "2. Avoid creating or reinforcing unfair bias." It states that

> AI algorithms and datasets can reflect, reinforce, or reduce unfair biases. We recognize that distinguishing fair from unfair biases is not always simple, and differs across cultures and societies. We will seek to avoid unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief.

While it can be acknowledged that efforts to address biases and unjust impacts on humans is welcomed, merely repeating legal requirements is not sufficient in the algorithmic age. Discrimination based on protected characteristics is prohibited in most jurisdictions around the world and should therefore clearly guide the company's actions. Amazon dedicates one paragraph to biases in its guide on Responsible Machine Learning[30] and utilizes a tool called Amazon

29 Meta, "Facebook's five pillars," *supra* note 26.

30 "Responsible Machine Learning," Amazon Web Services ("AWS"), accessed May 25, 2024, 2, https://d1.awsstatic.com/responsible-machine-learning/responsible-use-of-machine-learning-guide.pdf.

SageMakerClarify[31] to detect biases, the company defines biases as "imbalances in data or disparities in the performance of a model across different groups. Amazon SageMaker Clarify helps you mitigate bias by detecting potential bias during data preparation, after model training, and in your deployed model by examining specific attributes." Rather than merely stating policy goals, such examples present concrete tools that could be used by the company or as elements in a legislative framework to address the issues. Meta also disposes of tools to address biases, stating that Facebook "is continually improving our Fairness Flow tools and processes to help our ML engineers detect certain forms of potential statistical bias in certain types of AI models and labels."[32] These efforts are a notable step in the right direction, as they can serve as tools to achieve the objectives inside and outside legislative frameworks.

On *fairness*, Google acknowledges that

> First, ML models learn from existing data collected from the real world, and so an accurate model may learn or even amplify problematic pre-existing biases in the data based on race, gender, religion or other characteristics. For example, a job-matching system might learn to favor male candidates for CEO interviews, or assume female pronouns when translating words like "nurse" or "babysitter" into Spanish, because that matches historical data.

Discussing concrete examples of biases and potential discriminations is useful as it draws attention to the risks and shows the company's awareness of the problems. Microsoft, for example, set a goal to minimize stereotyping that states that "Microsoft AI systems that describe, depict, or otherwise represent people, cultures, or society are designed to minimize the potential for stereotyping (…) identified demographic groups, including marginalized groups." Facebook also states that: "To help us consider these issues from a broad range of perspectives, Facebook's Responsible Innovation team and Diversity, Equity & Inclusion team both facilitate input from a wide range of external experts and voices from underrepresented communities."[33] Here again, broad

31 "Amazon SageMaker Clarify," AWS, accessed May 25, 2024, https://aws.amazon.com/de/sagemaker/clarify/?sagemaker-data-wrangler-whats-new.sort-by=item.additionalFields.postDateTime&sagemaker-data-wrangler-whats-new.sort-order=desc.

32 Meta has even published an academic paper regarding the issue of fairness. See Chloé Bakalar et al., "Fairness On The Ground: Applying Algorithmic Fairness Approaches to Production Systems," arXiv, March 24, 2021, https://arxiv.org/abs/2103.06172.

33 Meta, "Facebook's five pillars," *supra* note 26.

TABLE 5.1 Human rights values contained in tech companies AI policies (human rights values modelled and based on largely on OECD 2020 recommendation on AI and values attributed on the basis of the analysis of the selected tech companies AI guidelines or principles)[a]

Human rights values	Amazon	Facebook	Google
Non-discrimination	(-)	(-)	(-)
Equality	(+)*	(+)*	(+)*
Diversity	(+)*	(+) *	(+)*
Fairness	(+) **	(+) **	(+) ***
Internationally recognized labour rights	(-)	(-)	(-)
Biases	(+) **	(+) **	(+) **

a (+) indicates that the company has a policy on the relevant human rights value and (-) indicates no dedicated policy was formulated in the AI guidelines. In addition, a value for each company policy is assigned in accordance with its relevance and suitability to address within the framework of non-binding guidelines the human rights values effectively with the relevant company policy (*** = adequate, ** = limited, * = not sufficient).

political statements are difficult to verify for consumers or regulatory authorities. Credibility would be added if accompanied by statistics or concrete examples of how the policy goals foster and impact the development of AI systems.

While the AI principles appear to be a marketing product in its own way – the wording is appealing and appears to be addressing a popular issue –, there are still many limits and shortcomings if companies aim to address and effectively diminish algorithmic discrimination. Notably, many company guidelines remain vague and, while they pose general principles or ideas, they do not concretely show how the company attempts to avoid discriminatory outcomes (see Table 5.1).

2.2 *Limits and Shortcomings*

In contrast to a national or regional legal standard, while most guidelines have many similarities, no single guideline addresses the same issues. This is true not only for the general issues of transparency, accountability, and responsibility, but also with regard to issues of equality and non-discrimination. This creates

different situations for consumers using a given service, as each business is free to set up their own particular guidelines which leads to legal uncertainty. Contrary to legal standards, an analysis of several guidelines also reveals the broad and imprecise formulation of some of these standards (see Table 5.1). Finally, while these guidelines can be invoked vis-à-vis the company itself and might eventually give rise to the ability to "litigate" rights within a system set up by the company,[34] they do not guarantee not guarantee the kind of gain cause that accompanies a lawsuit in a national court.

2.3 *Concrete Guidance Addressed at AI Developers*

Some non-binding documents, like the Fairness, Accountability, and Transparency in Machine Learning (FAT/ML), can be of added value, if companies use them while developing algorithms. For example, the Principles for Accountable Algorithms try to "Ensure that algorithmic decisions do not create discriminatory or unjust impacts when comparing across different demographics (e.g. race, sex, etc)." And the Social Impact Statement for Algorithms,[35] asks

> that algorithm creators develop a Social Impact Statement using the above principles as a guiding structure. This statement should be revisited and reassessed (at least) three times during the design and development process: design stage, pre-launch, and post-launch. When the system is launched, the statement should be made public as a form of transparency so that the public has expectations for social impact of the system.

Such non-binding principles can incorporate non-discrimination principles into algorithmic design. Using specific design questions templates is another effective approach to improve companies' impact statements. The method can be used by designers, developers, and tech companies during the model and

34 For example, see Meta, which set up the Facebook Oversight Boards which it wants to operate in a similar way as real courts of law where complaints can be launched, and decisions are published on the website. According to the company's explanation, "[t]he Oversight Board reviews content decisions made by Meta to see if the company acted in line with its policies, values, and human rights commitments. The Board can choose to overturn or uphold Meta's decision." See "Improving how Meta treats people and communities around the world," Oversight Board, accessed May 25, 2024, https://oversightboard.com.

35 Diakopoulos et al., "Principles for Accountable Algorithms," *supra* note 24.

design phase.[36] It can impact everything from the automated decision (Who is the audience and who will be most affected?), to bias detection (How to test the algorithm? What is the threshold for measuring and correcting biases notably for groups protected against discrimination by law), to diversity considerations (Is the design team representative enough?).[37] Such self-regulatory approaches can be used regardless of the existence of any current or future regulatory requirements and to mitigate bias and discrimination.

Specific guidance could take the form of a simple DIN-A4 page that is handed out to developers: it would recall some of the main risks of biases, stereotypes, and discrimination in relation to design of models, algorithms, and datasets. Even if the knowledge is already present in the developers' mindset or has been disseminated by training and seminars, it can make a difference to directly address the person in charge of developing the algorithm. Similar to the airline industry where pilots have been thoroughly trained, checklists are used to ensure the application of some fundamental principles and to ensure that key procedures are properly implemented. Such concrete guidance, specifically designed for those who create algorithms, can be made a requirement in legislative frameworks, which could support the reduction of biases and discriminatory potential at the design stage.

3 The Role of Business to Preserve Human Rights

Multi-stakeholder approaches to human rights are good practice.[38] One could argue that there is a special role and responsibility for businesses to preserve human rights[39] and avoid discriminatory outcomes in their algorithms. After

36 See, for example, Nicol Turner Lee, Paul Resnick, and Genie Barton, "Algorithmic bias detection and mitigation: Best practice and policies to reduce consumer harms," Brookings, May 22, 2019, https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/; specifically Table 1. Design questions template for bias impact statement.

37 See *supra* note 24, from which questions have been summarised.

38 See "Safeguarding freedom of expression and access to information: guidelines for a multistakeholder approach in the context of regulating digital platforms," United Nations Educational, Scientific and Cultural Organization ("UNESCO"), April 27, 2023, https://unesdoc.unesco.org/ark:/48223/pf0000384031.locale=en; Jonathan Andrew and Frédéric Bernard, *Human Rights Responsibilities in the Digital Age: States, Companies, and Individuals* (Bloomsbury, 2021).

39 See, for example, the United Nations Forum on Business and Human Rights, where the 12th Session in November 2023 is addressed "[t]owards effective change in implementing obligations, responsibilities and remedies," which also mentions "gender, business and

all, they create those AI systems which rather than being neutral or infallible,[40] have potentially biased and discriminatory effects on humans. Companies open and provide a platform, service, or product which can cause harm and violate human rights.[41] If a private business wants to make a profit by offering a service, it should also bear the responsibility to avoid and diminish human rights harms caused by the same AI technologies they profit from. In that sense, the recent Open Letter[42] signed by AI scientists and ethicists to temporarily ban AI technologies superior to Chat-GPT-4 is interesting because it raises the question of whether companies should have this moral responsibility or whether it should be incorporated into law.[43] In relation to the UNGPs,[44] a UN working group focusing on the issue of human rights and transnational corporations and other business enterprises recommended that businesses "invest in and sustain a focus on building their capacities to fulfill their responsibility to respect human rights across their activities, operations and business relationships."

It is argued here that throughout the of development AI systems – from the idea, conception of the model and the design of the algorithm to the deployment, monitoring, revision and correction of the algorithm – businesses should have obligations to ensure that they do not cause harm of a discriminatory nature to humans (3.1). Several proposed tools that are aimed at achieving less biases and discrimination will be analyzed from the lens of the role and the obligations of businesses (3.2).

---

human rights," as a standing issue of the Forum. "12th United Nations Forum on Business and Human Rights," OHCHR, November 27–29, 2023, https://www.ohchr.org/en/events/sessions/2023/12th-united-nations-forum-business-and-human-rights.

40 See Nicolas Sabouret and Laurent Bibard, *L'intelligence artificielle n'est pas une question technologique* (De l'aube, 2023), 39.

41 This logic also seems to underly the European Union's regulatory framework, such as the Digital Services Act. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance), OJ L 277, October 27, 2022.

42 "GPT-4 Technical Report," OpenAI, arXiv, March 4, 2024, https://arxiv.org/pdf/2303.08774.

43 Italy for example prohibited the use of Open AI's ChatGPT due to GDPR violations. The current Council of Europe legal proposal suggests the option for Member States to ban or temporarily limit the use of certain AI systems.

44 "Building capacity for the implementation of the Guiding Principles on Business and Human Rights: Report of the Working Group on the issue of human rights and transnational corporations and other business enterprises*," HRC, May 18, 2023, para. 84, https://www.ohchr.org/en/documents/thematic-reports/ahrc5324-building-capacity-implementation-guiding-principles-business.

3.1 *Business Responsibilities to Avoid Human Rights Harm and Discrimination throughout the Lifecycle of AI Systems*

Many guidelines, business consultancies, or legal standards contain references and links in their AI principles concerning the role of companies in the protection of human rights.[45] To avoid human rights violations, many companies actively foresee processes at company level.

The OECD Guidelines and Recommendations call on to respect the principle of non-discrimination and promote equality between women and men. The OECD Guidelines for Multinational Enterprises, for example, explicitly refer to non-discrimination in hiring practices as well as promotion practices (paras. 5), non-discrimination in employment and occupation (para. 51), the principle of non-discrimination, and refer to obligations contained in this regard in the International Labor Organization ("ILO") Conventions (para. 54).

The OECD/LEGAL/0449, Recommendation of the Council on Artificial Intelligence, which is understood to complement existing OECD instruments, includes recommendations both for governments and for AI actors, including businesses.[46] The OECD "CALLS ON all AI actors to promote and implement, according to their respective roles, the following Principles for responsible stewardship of trustworthy AI."[47]

Within the principle of human-centered values and fairness, the OECD states 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 labour rights."[48] On transparency and explainability, the Recommendation states that "AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art."

Under the principle of robustness, security and safety, it is recommended that "AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of

45 See for example Maria Luciana Axente and Ilana Golbin, "9 ethical AI principles for organizations to follow," World Economic Forum, June 23, 2021, https://www.weforum.org/agenda/2021/06/ethical-principles-for-ai/.

46 The G7 recently reaffirmed the OECD "Recommendation," *supra* note 8; see G7/G20, "Ministerial Declaration," *supra* note 9, at notably paras. 39–42.

47 OECD, "Recommendation," *supra* note 8, at Point III.

48 *Id*. at Point IV, 1.2 Human-centered values and fairness.

the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art" and that "AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias."

Finally, on accountability, it is recommended that "AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art."

These recommendations, addressed at businesses, are useful for regulators to consider and point to in order to identify the relevant areas where businesses can have a human rights impact. Notably, where businesses are the source of the information or the origin of the algorithm, it makes sense to reflect on the need for regulatory frameworks. UN reports also address recommendations to both states and businesses, but they tend to highlight some obligations for companies with particularity.

> The Toronto Declaration identifies three core elements or steps for corporate human rights diligence for machine learning systems: (a) identification of potential discriminatory outcomes; (b) prevention and mitigation of discrimination and tracking of responses; and (c) transparency regarding efforts to identify, prevent and mitigate discrimination.[49]

Similarly, the Council of Europe originally assigned some obligations to states and others to businesses in the proposed Draft Framework Convention on AI. While Chapter II and specifically Articles 5–7 addressed public authorities, Article 8 for example specifically addressed private actors.[50] However, in the final version of the Framework Convention, the scope is restricted to public

---

49 Toronto Declaration, *supra* note 10, at paras. 44–51, as summarized in HRC, "Racial discrimination," *supra* note 16, at para. 60.

50 Article 8(b)of the Draft Convention: "effective guidance is provided to relevant public and private actors on how to prevent and mitigate any adverse impacts of the application of an artificial intelligence system on the enjoyment of human rights and fundamental freedoms, the functioning of democracy and the observance of the rule of law in their operations." "Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law," Committee on Artificial Intelligence ("CAI"), January 6, 2023, Article 8(b), https://rm.coe.int/cai-2023-01-revised-zero-draft-framework-convention-public/1680aa193f.

actors and includes only private actors acting on behalf of public authorities (Art. 3(1)(a)).

While the European Union's AI Act, adopted by co-legislators in 2024, takes the legal nature of an EU Regulation, it assigns obligations to companies, states and supervisory authorities, but in particular to the developers and users of AI systems that want to use systems within the EU market or where the impacts of the AI system affect the EU market. In that sense, the EU AI Act can be seen as a sort of product regulation which produces compliance requirements and costs for companies.

Major AI companies use tools throughout the AI lifecycle to ensure fairness and/or detect biases in their algorithms or machine learning tools.[51] Other resources used by AI companies include specific guides for AI developers and programmers as well as education and training[52] to ensure fair and non-biased algorithms. It would be wise for global AI regulators to take note of the existing assessment infrastructure and capabilities among AI companies when they design new rules. They should also work with independent researchers to assess the concrete needs and feasibility of regulatory obligations, for example to detect, diminish, and eliminate biases from the design, training and other datasets in order to reduce discriminatory outcomes.

3.2 *The Tools of the Soft Law Instruments and Proposed Regulatory Frameworks to Achieve Non-discriminatory AI Systems*

Among the soft law instruments, the tools suggested can be regrouped in three categories based on the amount of involvement required by businesses and potential compliance efforts: information, active and passive obligations, and far-reaching involvement. Information requirements include providing information to AI systems users and publishing relevant information such as reports on bias audits on websites. Active obligations include conducting ex ante assessments, such as bias audits, to check algorithms for potential biased datasets, biased design, or discriminatory impacts. Passive requirements include all regulatory actions by public authorities that check or verify compliance with

51 "How we're using Fairness Flow to help build AI that works better for everyone," Meta, March 31, 2021, https://ai.facebook.com/blog/how-were-using-fairness-flow-to-help-build-ai-that-works-better-for-everyone/; AWS, "Amazon SageMaker Clarify," *supra* note 31.

52 Amazon for example considers that "Continuous education on the latest developments in ML is an important part of responsible use. AWS offers the latest in ML education across your learning journey through programs like the AWS Machine Learning University (Bias and Fairness Course)." These videos are available online at Machine Learning University, "Responsible AI," YouTube, November 22, 2022, https://www.youtube.com/playlist?list=PL8P_Z6C4GcuVMxhwT9JO_nKuWOQMSJ-cZ.

a given regulation and granting access to businesses for certain activities, like sourcing code or creating datasets.

Among the proposed hard law instruments, several tools are usually proposed to ensure that regulatory goals like bias mitigation and reducing discriminatory impacts of algorithms are met. First, several proposed international frameworks[53] suggest bias audits, a tool that can be used either before market entrance or for monitoring the AI system while in use. Typically, the aim of a bias audit is to detect any biases or discriminatory outcomes and correct them before releasing an AI system on the market. Second, known from environmental impact or human rights, algorithmic impact assessments are considered a comprehensive tool used to check for compliance with bias and discrimination requirements. Companies need to provide detailed information to show that they made all required efforts to ensure no or less biases and discrimination in their algorithms.[54] Third, transparency and documentation tools enable potential victims of discrimination and authorities to verify any violations. Fourth, prohibition is a strong tool used to prohibit the use of an AI system in very specific circumstances in an effort to avoid human rights harms. Fifth, an essential tool to ensure compliance is the possibility to impose fines on companies that develop or use AI systems.

4 Legislative Human Rights Frameworks for AI (Hard Law)

Human Rights frameworks that apply business expectations in differing degrees are found at UN level, notably in several reports outlining recommendations for businesses and states (4.1). At the Council of Europe (CoE) level, a legal framework on AI and Human Rights is in the works (4.2), and at the EU level, the AI Act has been just adopted, which is based on the respect of fundamental rights (4.3). While both the CoE and the EU legislative frameworks will be enforceable by national and the Luxembourg and Strasbourg courts, the proposed ideas at UN level are less binding but have been nevertheless placed in this section due to their potential repercussions. Regardless of the type of

53 See for example, the NYC Law, *infra* TABLE 5.2, The White House, "Blueprint," *supra* note 15, the EU AI Act, *infra* TABLE 5.2, but also mentioned as options in the OECD, "Recommendation," *supra* note 8, and the CoE Framework Convention, *infra* TABLE 5.2.

54 See for example the "Algorithmic Impact Assessment tool," Government of Canada, accessed May 25, 2024, https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html.

TABLE 5.2 Tools to achieve non-discriminatory AI systems[a]

Tools	Soft law	Hard law	Proposals
Risk management system or Bias audits to address risk of biases	OECD (1.4.c)	NYC (§ 20–870, 20–871)	EU (Art. 9), CoE (Articles 16, 12, 24)
Human Rights Impact Assessments (HRIA) or Algorithmic Impacts Assessment (AIA)	(-)	(-)	EU (Art. 27)[b], CoE (Art. 16)
Transparency and Documentation tools	OECD (1.3)	NYC (§ 20–871)	EU (Art. 10–13), CoE (Art. 8)
Remedies/ Complaint mechanism	OECD (1.5)	NYC (§ 20–872)	EU (Art. 85, 86), CoE (Art. 14, 15)
Fines/Penalties	OECD (1.5)	NYC (§ 20–872)	EU (Art. 97)

a This table analyses the tools that are mentioned in the existing soft (OECD, CoE) and hard law frameworks (A Local Law to amend the administrative code of the city of New York, in relation to automated employment decision tools, Law 2021/144, December 11, 2021), text available online at https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=ID%7CText%7C&Search= ("NYC Law") as well proposed legislative frameworks on AI ("Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law," CoE, adopted May 17, 2024, https://rm.coe.int/1680afae3c ("CoE Framework Convention"); Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM(2021) 206 final, 2021/0106(COD), April 21, 2021 ("EU AI Act")). The final text, as adopted by co-legislators was published on May 21, 2024, text available online at https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/en/pdf ("AI Act"). While the AI Act will in principle apply 24 months after entry into force of the Regulation, some specific provisions apply after 6 months, 12 months or 36 months, Art. 113 AI Act.

b Spain adopted a Digital Rights Charter in 2021, which includes specific principles on equality and non-discrimination; see notably La Moncloa, "Carta Derechos Digitales," *supra* note 14, at Article VIII.

TABLE 5.3 Legislative human rights frameworks

	United Nations	Council of Europe	European Union
General AI framework	(-) Currently, only reports and political calls	(+)	(+) but among the general principles, equality and non-discrimination
Discrimination-specific instrument	(+) CEDAW as general instrument and currently a General Recommendation (N°40 is drafted which includes considerations on AI and algorithmic discrimination)	(-) but planned Legal Instrument by 2025	(-) partially some of the issues of biases and non-discrimination could be addressed with the regulation of AI systems in the area of education and labor market (Ex:AI recruitment systems)
Equality and Discrimination principles	(+) UNGPS	(+)	(+)
Specific requirements or recommendations for business	(+)	(+) Article 8	(+) The requirements are mostly addressed to businesses

instrument or the level of implementation, the impact from the point of view of human rights law is key.[55]

55 See for example Matilda Arvidsson and Gregor Noll, "Artificial Intelligence, Decision Making and International Law," *Nordic Journal of International Law* 92, no. 1 (April 2023): 1–8.

4.1 *UN Level*

The Report of the Special Representative of the Secretary-General John Ruggie on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises, or the "Guiding Principles on Business and Human Rights" ("UNGPs"), is a major reference document for business and human rights.[56] The UNGPs, adopted by the UN Human Rights Council by Resolution 17/4 on 16 June 2011, are a set of principles directed at governments and businesses that clarify their duties and responsibilities in the context of business operations. For instance, Pillar 2 spells out foundational and operational principles that specify businesses' responsibility to avoid social impacts wherever they operate and whatever their size or industry, and address any impact that does occur. The UNGPs rely on three pillars focused on business' obligation to respect, states' obligation to protect and the possibility for accountability and remedies.[57]

Principle 11 stipulates in general that "Business (…) should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved." Principle 12 further specifies what that means in terms of the legal framework: "The responsibility of business enterprises to respect human rights refers to internationally recognized human rights – understood, at a minimum, as those expressed in the International Bill of Human Rights and the principles concerning fundamental rights set out in the International labour Organization's Declaration on Fundamental Principles and Rights at Work." Principle 13 then details specific obligations of businesses:

> The responsibility to respect human rights requires that business enterprises: (a) Avoid causing or contributing to adverse human rights impacts through their own activities, and address such impacts when they occur; (b) Seek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts.

56 HRC, "UNGPs," *supra* note 16.

57 For a specific application of the UNGPs, which have been designed in a world before algorithms and AI, see the specific context of AI in the "B-Tech Project: Multi-Stakeholder Consultation on Gender, Tech, and the Role of Business," OHCHR, June 15, 2023, https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/B-Tech-gender-multi-stakeholder-consultation.pdf.

The operational principles clarify the general human rights commitments and show ways in which these can be implemented at company level, for example through policy statements,[58] Human Rights Due Diligence,[59] assessment and identification of actual or future human rights risks,[60] prevention and mitigation of adverse human rights impacts,[61] monitoring and tracking of human rights response,[62] and external communication of human rights efforts.[63] Considering that the UNGPs have been designed with both businesses and governments in mind, they incorporate many concrete steps, mechanisms, and procedures to mitigate human rights impacts which are also applicable in the context of algorithmic discrimination. Thus, they can inspire the drafting of national and regional legal frameworks to address algorithmic discrimination. The operational principles are particularly well suited to achieve this end. While policy statements might be less adequate to mitigate human rights harm, human rights due diligence and the assessment and identification of actual or future human rights harm are suitable tools to prevent and mitigate algorithmic discrimination.[64]

The Report of the Special Rapporteur on the rights of persons with disabilities[65] is one of the rare UN reports that addressed issues of AI and is therefore of relevance in this context. Three aspects are highlighted in the Report's conclusions that are relevant to the current analysis. First, the "unprecedented power of artificial intelligence [which] can be a force for good for persons with disabilities," and that the "Profound advances for humankind must be properly harnessed to make sure that the farthest left behind can at last benefit fully from science and its advancements." Second, it acknowledges that "the well documented negative impacts of artificial intelligence on persons with disabilities need to be openly acknowledged and rectified by states, business, national human rights institutions, civil society and organizations of persons with disabilities working together." In accordance with the present analysis, it pointed out that

58 UNGPs, *supra* note 16, at Principle 16.
59 *Id*. at Principle 17.
60 *Id*. at Principle 18.
61 *Id*. at Principle 19.
62 *Id*. at Principle 20.
63 *Id*. at Principle 21.
64 See for example the Proposal for a Directive of the European Parliament and of the Council on Corporate Sustainability Due Diligence and amending Directive (EU) 2019/1937, COM/2022/71 final, 2022/0051(COD), February 23, 2022.
65 HRC, "Rights of persons with disabilities," *supra* note 16.

> At the development level, those negative impacts arise from poor or unrepresentative data sets that are almost bound to lead to discrimination, a lack of transparency in the technology (making it nearly impossible to reveal a discriminatory impact), a short-circuiting of the obligation of reasonable accommodation, which further disadvantages the disabled person, and a lack of effective remedies. While some solutions will be easy and others less straightforward, a common commitment is needed to work in partnership to get the best from the new technology and avoid the worst.

Third, and finally, the document calls for "a fundamental reset of the debate (…) based on more evidence and greater consideration of the rights and obligations contained in the Convention on the Rights of Persons with Disabilities and other human rights instruments." The recommendations specifically address businesses and the private sector,[66] notably in relation to transparency and information obligations,[67] disability-inclusive human rights impact assessments for AI,[68] human rights due diligence,[69] accessible and effective non-judicial remedies and redress for AI caused human rights harms,[70] and realistic and representative datasets.[71]

---

66 *Id*. at para. 78.

67 *Id*. at para. 78(a): "[o]perate with transparency and provide information about how artificial intelligence systems work. That should include alignment with open-source and open data standards and publication of accessible information about how artificial intelligence systems operate."

68 *Id*. at para. 78(b): "[i]mplement disability-inclusive human rights impact assessments of artificial intelligence to identify and rectify its negative impacts on the rights of persons with disabilities. All new artificial intelligence tools should undergo such assessments from a disability rights perspective. Artificial intelligence businesses should conduct their impact assessments in close consultation with organizations representing persons with disabilities and users with disabilities."

69 *Id*. at para. 78(c): "[u]se corporate human rights due diligence to explicitly take account of disability and artificial intelligence. Private sector actors that develop and implement machine-learning technologies must undertake corporate human rights due diligence to proactively identify and manage potential and actual human rights impacts on persons with disabilities, to prevent and mitigate known risk in any future development."

70 *Id*. at para. 78(d): "[e]nsure accessible and effective non-judicial remedies and redress for human rights harms arising from the adverse impacts of artificial intelligence systems on persons with disabilities. This should complement existing legal remedies and align with the International Principles and Guidelines on Access to Justice for Persons with Disabilities."

71 *Id*. at para. 78(d): "[e]nsure that data sets become much more realistic and representative of the diversity of disability and actively consult persons with disabilities and

In the UN Report on Racial discrimination and emerging digital technologies[72] racial discrimination is analyzed from the perspective of AI. The Report analyzes different forms of racial discrimination in the design and use of emerging digital technologies, including the structural and institutional dimensions of this discrimination and outlines both human rights obligations of states and the responsibility of businesses to combat this discrimination. The report highlights the "monumental influence in the design and use of emerging digital technologies"[73] of private corporations but also emphasizes them as "key intermediaries between governments and their nations, with the capacity to significantly transform the situation of human rights."[74] It is recalled that human rights law "will by no means be a panacea for the problems identified (…) but it stands to play an important role in identifying and addressing the social harms of artificial intelligence and ensuring accountability for these harms."[75] In addition, the report emphasizes that "[t]hese obligations also have implications for non-State actors, such as technology corporations, which in many respects exert more control over these technologies than states do."[76] With regard to the diversity crisis discussed earlier (Section 2), the Report mentions states' obligations to prevent and eliminate racial discrimination in the design and use of AI but stresses that "states must work together with private corporations"[77] to ensure meaningful diversity and equal representation and suggests including opportunities for "co-design and co-implementation with representatives of racially or ethnically marginalized groups"[78] when it comes to conducting impact assessments. In this regard, the Report recalls that human rights law is only legally binding on states, who in turn must ensure that effective remedies for racial discrimination attributable to private actors, including corporations, exist in national or regional laws.[79] The Business and Human Rights in Technology Project (B-Tech Project)

---

their representative organizations when building technical solutions from the earliest moments in the business cycle. This includes proactively hiring developers of artificial intelligence who have lived experience of disability, or consulting with organizations of persons with disabilities to gain the necessary perspective."

72 HRC, "Racial discrimination," *supra* note 16.

73 *Id*. at para. 15.

74 *Id*. at para. 16.

75 *Id*. at para. 45.

76 *Id*.

77 *Id*. at para. 55.

78 *Id*. at para. 56.

79 *Id*. at para. 59.

is mentioned as an example "which applies the Guiding Principles to digital technologies, due diligence should apply to the conceptualization, design and testing phases of new products – as well as the underlying data sets and algorithms that support them."[80]

The recommendations notably state that

> States must ensure that human rights ethical frameworks for corporations involved in emerging digital technologies are linked with and informed by binding international human rights law obligations, including on equality and non-discrimination. There is a genuine risk that corporations will reference human rights liberally for the public relations benefits of being seen to be ethical, even in the absence of meaningful interventions to operationalize human rights principles. Although references to human rights, and even to equality and non-discrimination, proliferate in corporate governance documents, these references alone do not ensure accountability. Similarly, implementation of the framework of Guiding Principles on Business and Human Rights, including through initiatives such as the B-Tech Project, must incorporate legally binding obligations to prohibit – and provide effective remedies for – racial discrimination.[81]

In addition, the report identifies the main concerns with non-binding guidelines:

> An inherent problem with the ethics-based approaches that are promulgated by technology companies is that ethical commitments have little measurable effect on software development practices if they are not directly tied to structures of accountability in the workplace. From a human rights perspective, relying on companies to regulate themselves is a mistake, and an abdication of State responsibility. The incentives for corporations to meaningfully protect human rights (especially for marginalized groups, which are not commercially dominant) can stand in direct opposition to profit motives. When the stakes are high, fiduciary obligations to shareholders will tend to matter more than considerations

80 *Id*. at para. 60; See also "Business and Human Rights in Technology Project ("B-Tech Project"): Applying the UN Guiding Principles on Business and Human Rights to Digital Technologies," OHCHR, accessed May 25, 2024, https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/BTechprojectoverview.pdf.

81 HRC, "Racial discrimination," *supra* note 16, at para. 61.

> concerning the dignity and human rights of groups that have no means of holding these corporations to account. Furthermore, even well-intentioned corporations are at risk of developing and applying ethical guidelines using a largely technological lens, as opposed to the broader society-wide, dignity-based lens of the human rights framework.[82]

Finally, the report suggests that corporate human rights due diligence needs to be implemented by states based on human rights law prohibitions on racial discrimination and refers to the European Commission's proposal for mandatory due diligence for companies.[83]

The UN Report on the right to privacy in the digital age[84] has also outlined several obligations and recommendations for businesses in the context of AI and algorithmic discrimination, notably with regard to gender equality and non-discrimination. The report analyzes how widespread AI use by states and businesses affects the enjoyment of the right to privacy and associated rights, identifies the state-business nexus[85] and provides some recommendations for states and businesses regarding the design and implementation of safeguards to prevent and minimize harmful outcomes and to facilitate the full enjoyment of the benefits that artificial intelligence can provide. The Special Rapporteur recommends business enterprises to:

> (a) Make all efforts to meet their responsibility to respect all human rights, including through the full operationalization of the Guiding Principles on Business and Human Rights;
> (b) Enhance their efforts to combat discrimination linked to their development, sale or operation of AI systems, including by conducting systematic assessments and monitoring of the outputs of AI systems and of the impacts of their deployment;
> (c) Take decisive steps in order to ensure the diversity of the workforce responsible for the development of AI;
> (d) Provide for or cooperate in remediation through legitimate processes where they have caused or contributed to adverse human

82 *Id*. at para. 62.

83 *Id*. at para. 63; "European Commission Promises Mandatory Due Diligence Legislation in 2021," Responsible Business Conduct ("RBC"), April 30, 2020, https://responsiblebusinessconduct.eu/wp/2020/04/30/european-commission-promises-mandatory-due-diligence-legislation-in-2021.

84 HRC, "The right to privacy in the digital age," *supra* note 16.

85 *Id*. at paras. 51–54.

> rights impacts, including through effective operational-level grievance mechanisms.[86]

The two most recent work-streams at the UN level concern the Commission on the Status of Women ("CSW"). Each year, the CSW adopts agreed conclusions as a result of the 67th session which was dedicated in 2023 mainly to digital and AI topics in the context of women's rights.[87] In addition, the current work of the Committee on the elimination of all discriminations against women ("CEDAW committee") regarding the elaboration of a General Recommendation 40[88] on equal representation in decision making systems is recognizing AI systems as game-changing technology.

4.2 *Council of Europe Level*

The Council of Europe ("CoE") is a major force that sees itself as the guardian of human rights and democracy. Equality and non-discrimination have been on its agenda and its core instruments since the beginning. The CoE adopted a Framework Convention on AI and Human Rights, Democracy and the Rule of Law[89] in May 2024 and is currently working towards a framework for AI regulation that specifically addresses equality, including gender equality, and non-discrimination angle which will be published by 2025.[90] The clear advantage of CoE legal frameworks on AI are deeply rooted in human rights, enforceable by the European Court of Human Rights in Strasbourg, and has a wide potential reach within the CoE's 46 Member States. In addition, other states are following the CoE regulatory process on AI carefully as "representatives

86 *Id*. at para. 61.

87 See "CSW67 (2023)," United Nations Women, March 6–17, 2023, https://www.unwomen.org/en/csw/csw67-2023.

88 See Fabian Lütz, "Written submission (focusing on AI, automated decision-making systems and gender equality) for the half-day General Discussion on the equal and inclusive representation of women in decision-making systems, 84th session of CEDAW," OHCHR, February 22, 2023, https://www.ohchr.org/sites/default/files/documents/hrbodies/cedaw/general-discussion/2023/hdgd-20230222-luetz-equality-representation-ai.docx; see also Tetyana (Tanya) Krupiy, "Meeting the Chimera: How the CEDAW Can Address Digital Discrimination," *International Human Rights Law Review* 10 (June 2021): 1–39.

89 CAI, "Revised Zero Draft," *supra* note 50.

90 *Id*.; "Council of Europe's Work in progress," Council of Europe, last updated January 2024, https://www.coe.int/en/web/artificial-intelligence/work-in-progress.

of the observer states," such as Canada, Israel, Japan, and the United States of America which all have many AI companies within their jurisdictions.[91]

First, the Framework Convention on AI includes rules on the principle of non-discrimination (Art. 10). Previously, Article 12 specifically stated that "Each Party shall (…) ensure that the design, development and application of artificial intelligence systems respect the principle of equality, including gender equality and rights related to discriminated groups and individuals in vulnerable situations." Now only the Preamble makes reference to "the risks of discrimination in digital contexts, particularly those involving artificial intelligence systems, and their potential effect of creating or aggravating inequalities, including those experienced by women and individuals in vulnerable situations."

Second, the equality and non-discrimination legal instrument is too far in the future to make specific remarks, but it will be surely based on and incorporate the work of the general framework Convention and consider the work of the EU AI Act.

4.3 *European Union Level*

The EU is currently the most advanced jurisdiction in the world in terms of addressing adverse effects of AI and algorithms on fundamental rights. Following the EU General Data Protection Regulation ("GDPR"), the EU adopted the Digital Services Act ("DSA") and the Digital Markets Act ("DMA") and set up the European Center for Algorithmic Transparency ("ECAT"), tasked with helping the European Commission enforce the DSA with relevant expert knowledge.[92] Considering that in principle, all EU legislation is based on a human and fundamental rights approach, EU law can be an interesting reference point because it offers an example of legally binding nature and model character that is often imitated by other jurisdictions.[93]

The most important EU legislation in the area of AI, the AI Act, was adopted by both co-legislators (European Parliament in March 2024 and Council of the European Union in May 2024). The AI Act is a legally binding instrument for

91 See the members and observers of the CAI, located at "Committee on Artificial Intelligence (CAI)," accessed May 25, 2024, https://www.coe.int/en/web/artificial-intelligence/cai#{%22126720142%22:[0]}.

92 "European Centre for Algorithmic Transparency," European Commission, accessed May 25, 2024, https://algorithmic-transparency.ec.europa.eu/index_en.

93 See notably, Anu Bradford, "The Brussels Effect," *Northwestern University Law Review* 107, no. 1 (December 2012): 1–68; Anu Bradford, "Chapter 9: The Future of the Brussels Effect," in Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press, 2020).

regulating AI on the basis of a fundamental rights approach which foresees several very specific obligations for AI companies designing or using AI systems. In that regard, it differs from the previously discussed soft law instruments because, far from being broad and political, the relevant articles detail the legal requirements imposed on companies.[94] The AI Act also involves private stakeholders, via the specification of technical standards. It allows for the European Standard Setting Organizations to specify requirements for certain concepts and mechanisms used in the legislative framework of the AI Act.[95]

Specifically, on algorithmic discrimination, the AI Act could make a fundamental contribution to controlling and diminishing discrimination caused by AI systems. While not obvious at first sight, the horizontal proposal on AI regulation could address the issue of discrimination through the regulatory requirements for High-risk AI systems. The scope of High-risk AI systems would include areas such as education or labor market use cases, for example AI recruitment systems.[96] If an AI recruitment system is classified as High-risk, it would need to comply with the requirements, like mandating a certain level of transparency and documentation. As such, if AI applications fall under the scope of the Directive, they would need to fulfill the detailed requirements for High-risk AI systems. This would enable control of the design and datasets of those AI systems, which in turn could help identify and diminish the risk of discriminatory outcomes. The European Parliament wanted to include the principle of discrimination in the operative part of the legislative text.[97]

---

94 This chapter does not discuss the substantive content of the AI Act in detail. For a summary and more information see notably, Fabian Lütz, "Gender equality and artificial intelligence in Europe. Addressing direct and indirect impacts of algorithms on gender-based discrimination," *ERA Forum* 23 (April 2022): 33–52, https://doi.org/10.1007/s12027-022-00709-6.

95 See "Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence," European Commission, December 5, 2022, https://ec.europa.eu/docsroom/documents/52376.

96 See EU AI Act, *supra* TABLE 5.2, at Annex III. In the US, specific guidance was recently issued by the U.S. Equal Employment Opportunity Commission ("EEOC") in relation to algorithmic recruitment and disability; see The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees, EEOC-NVTA-2022-2, May 12, 2022 (text available online at https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence) ("The ADA and AI: Applicants and Employees").

97 The European Parliament proposed to include an article 4a – General principles applicable to all AI systems e) which reads "diversity, non-discrimination and fairness' means that AI systems shall be developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law." See the

This would have been an improvement compared to the original European Commission proposal which merely foresaw the issue of biases and discrimination by AI systems in the recitals.[98] In addition, the European Parliament suggested strengthening the enforcement mechanisms of the AI Act, notably by including a complaint mechanism (which has been integrated in Art. 85),[99] upgrading the EU AI Office in the future to the status of an EU Agency (not reflected in the AI Act),[100] foreseeing joint investigations between national and EU authorities (now Art. 74 (11))[101] and strengthening the access to the models and algorithms underlying AI systems (now for example included inter alia in Art. 77).[102] These suggestions would help detect algorithmic discrimination and facilitate claims by victims of discrimination. While it is clear that the aim of the EU AI Act is not to address the issue of algorithmic discrimination *per se*, it can be a useful complement to classical non-discrimination laws and can facilitate detection and enforcement of non-discrimination claims as specified. Because it only partially addresses potential problems of biases and discrimination in relation to the AI applications, there is room to fully address the issue of algorithmic discrimination in any future review of the EU Gender Equality and Non-discrimination Directives.[103] Together with other future legislative frameworks, such as the proposed EU AI Liability Directive,[104] enforcing non-discrimination claims in the age of AI might be strengthened, notably when it comes to available remedies and compensations for violations of EU

adopted text of the European Parliament in its first reading: "P9_TA(2023)0236: Artificial Intelligence Act – Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) (Ordinary legislative procedure: first reading)," European Parliament, June 14, 2023, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf.

98 See EU AI Act, *supra* TABLE 5.2, at recital 36.

99 *Id*. at Article 68a and Recital 84a.

100 *Id*. at Article 56.

101 *Id*. at Article 66a.

102 *Id*. at Article 64.

103 For some proposals and recommendations specifically on EU law, see Fabian Lütz, "Algorithmische Entscheidungsfindung aus der Gleichstellungsperspektive – ein Balanceakt zwischen Gender Data Gap, Gender Bias, Machine Bias und Regulierung," *GENDER – Zeitschrift für Geschlecht, Kultur und Gesellschaft* 15, no. 1 (2023): 26–41, www.budrich-journals.de/index.php/gender/article/view/41777; Fabian Lütz, "Gender equality and artificial intelligence in Europe," *supra* note 97.

104 Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM(2022) 496 final, 2022/0303(COD), September 28, 2022.

non-discrimination rules in conjunction with the liability principles of the future EU AI Liability Directive. A joint legislative framework that includes elements of *ex ante* regulation of AI systems, reformed non-discrimination laws to consider algorithmic discrimination[105] as well as AI liability frameworks that would be applicable to remedies for non-discriminatory AI systems would enable victims of human rights violations and discriminations to have effective remedies.

5 Avoiding a Shift from Classical Public Lawmaking towards Private Rule-Setting: Advantages and Limits of Including Businesses in Private Regulatory Tasks to Avoid Discrimination

In the age of algorithms, the fast speed with which AI companies develop new algorithms and AI tools coincides with, or triggers, attempts to shape the global AI landscape, which is currently mostly unregulated.[106] This facilitates the risk of *de facto* private rules that govern AI systems in the absence of hard laws. AI companies develop their products in an AI race with regulators running behind in a regulatory race in order to include the latest AI inventions into their regulatory design and ensuring that the AI definition used and the regulatory tools included in the proposed legislative frameworks still adequately

105 The elections of the European Parliament in 2024 and the newly composed European Commission in terms of the College of Commissioners might be an opportunity to include a reform of the gender equality and non-discrimination Directives on the agenda and in the Commission Work Programme in order to fully take into account the effects and impacts of algorithmic discrimination in the EU acquis in order to complement the legal framework on AI composed most likely by the soon-to-be adopted EU AI Act. Notably, 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), OJ L 204, July 26, 2006, and 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, OJ L 373, December 21, 2004, seem good candidates to be reviewed in line of algorithmic discrimination.

106 See in general, Paul Nemitz, "Constitutional democracy and technology in the age of artificial intelligence," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2133 (October 2018); Paul Nemitz and Eike Gräf, "Artificial Intelligence Must Be Used According to the Law, or Not at All," Verfassungsblog: On Matters Constitutional, March 28, 2022, https://verfassungsblog.de/roa-artificial-intelligence-must-be-used-according-to-the-law/.

capture the newest AI systems.[107] In light of those challenges, involving businesses in the regulatory process embodies both advantages (5.1) and some limits (5.2) which need to be carefully balanced in order to achieve the regulatory goal of avoiding or reducing discrimination by AI systems.

5.1 *Advantages of Involving Businesses: Expertise*

There are several clear advantages to involving businesses in the fight against discrimination from the start, including any positive changes to the conception of models, design of algorithms, and AI systems to their deployment, and the post-market monitoring. The added value that is provided by the AI businesses and their researchers can take the form of pure knowledge (sharing) or amount to a gradual involvement in development and regulation.

First, and foremost, the developers of AI systems that program the underlying algorithms have the knowledge of what views and concepts shaped their design of the algorithm and how it functions. Sharing this knowledge with regulators enables the latter to make informed decisions more easily, considering that some relevant information is only held by the companies. Often, regulators depend on the flow of information from businesses to administrations to enforce the rules. While the state can impose access to information under specific circumstances, voluntary collaboration is preferred and less costly.

Second, including AI businesses from the outset – when algorithms are being designed – could help to incorporate the principle of non-discrimination *by design*.[108]

Third, soft law and hard law proposals often prescribe specific obligations and responsibilities for companies, ranging from transparency, information, and documentation requirements to the obligation to conduct AI impact assessments,[109] audits, or monitoring. In each these scenarios, a close

107 See the example of the EU AI Act, *supra* TABLE 5.2, which dates from 2021 and which did not fully consider AI systems such as Large Language Models (LLMs), and that Member States and the European Parliament now call for LLMs to be included in the regulatory efforts.

108 Tilburg University, "Non-discrimination by design," Tilburg University, 2019, https://www.tilburguniversity.edu/sites/default/files/download/04%20handbook%20non-discrimination%20by%20design%28ENG%29.pdf; Jonas Rebstadt et al., "Non-Discrimination-by-Design: Handlungsempfehlungen für die Entwicklung von vertrauenswürdigen KI-Services," *HMD Praxis der Wirtschaftsinformatik* 59, no. 2 (March 2022): 495–511, https://doi.org/10.1365/s40702-022-00847-y.

109 Some companies foresee Guides ("Microsoft Responsible AI Impact Assessment Guide," Microsoft, June 2022, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4ZzOI) and templates on how to conduct these AI Impact Assessments, allegedly in order to "contribute to the discussion about building better norms and practices around AI;" see,

collaboration between the state and businesses is essential. Only if businesses are involved from the start of the legislative process and provide information about the feasibility of specific regulatory ideas, can regulation can be designed appropriately and effectively implemented. While there might be differences of opinion to the extent and depth of regulation, early collaboration can ensure that both actors are aware of the regulatory environment which helps companies adopt the future regulation in terms of compliance and helps states to better enforce their new regulatory framework.

Fourth, regulation can envisage a specific role for business in the development and implementation of regulatory content. Business representatives might be present in oversight bodies or be entrusted with specific roles or responsibilities to help shape the legal rules.[110] An example can be found in the proposed EU AI Act, which foresees the specification of some of the content of the future Regulation within the framework of a standardization request entrusted to European Standard setting organizations, the European Committee for Standardization ("CEN") and the European Electrotechnical Committee for Standardization ("CENELEC").[111] Nevertheless, considering the risk of regulatory capture, the state needs to be aware of potential imbalances of knowledge and technical expertise, which could lead to a situation where states have to blindly follow the analysis or assessment of companies due to a lack of expertise on the administration's side.[112]

---

for example, "Microsoft Responsible AI Impact Assessment Template," Microsoft, June 2022, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFk.

110 Pieter Van Cleynenbreugel, "EU By-Design Regulation in the Algorithmic Society: A Promising Way Forward or Constitutional Nightmare in the Making?," in *Constitutional Challenges in the Algorithmic Society*, eds. Hans-W. Micklitz et al. (Cambridge University Press, 2021).

111 "A Notification under Article 12 of Regulation (EU) No 1025/2012: Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence," European Commission, December 5, 2022, https://artificialintelligenceact.eu/wp-content/uploads/2022/12/AIA-%E2%80%93-COM-%E2%80%93-Draft-Standardisation-Request-5-December-2022.pdf.

112 See in this regard recent efforts in the U.S. to gain further understanding of AI systems from the AI industry, "FACT SHEET: Biden-Harris Administration Takes New Steps to Advance Responsible Artificial Intelligence Research, Development, and Deployment," The White House, May 23, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/23/fact-sheet-biden-harris-administration-takes-new-steps-to-advance-responsible-artificial-intelligence-research-development-and-deployment/; The White House launched a Request for Information from the public, "Request for Information: National Priorities for Artificial Intelligence," The White House, May 23, 2023, https://www.whitehouse.gov/wp-content/uploads/2023/05/OSTP-Request-for-Information-National-Priorities-for-Artificial-Intelligence.pdf; in August 2023, some AI companies agreed to undergo a public assessment of their generative AI models; see Makenzie Holland, "Top tech firms

While all these various roles and types of involvement can have beneficial impacts for good and effective design and implementation of regulation, there are some important limits that should be considered when involving business in the regulatory process.

5.2 *Limits of Involving Businesses in the Regulation of AI: Lack of Legitimacy and Private Interests*

First, business decisions are usually steered by private and economic considerations rather than common good considerations. The businesses' interests might be in conflict with citizens interests' or the state's interest. Businesses are not democratically legitimized and accountable in the same way as public bodies.

Second, lack of resources on the side of government institutions might create a temptation to over-rely on business inputs and expertise without being able to understand or duly verify its impacts.

Third, independence can be a problem even if a lot of research is of high quality and standards.[113] In addition, high salaries facilitate talent to be attracted by business rather than the resource-drained public administrations or academia, which further concentrates knowledge and power on AI systems within the industry[114] on which often regulators can only rely because they lack the means and the information. Even though academic knowledge can also foster ideas, review, and improve company processes[115] to preserve human rights and the principle of non-discrimination, the ultimate aim of each company employee remains to work in the interest of the company.

---

agree to help White House probe AI risks," TechTarget, May 4, 2023, https://www.techtarget.com/searchcio/news/366536400/Top-tech-firms-agree-to-help-White-House-probe-AI-risks.

113 Roman Jurowetzki et al., "The Privatization of AI Research(-ers): Causes and Potential Consequences – From university-industry interaction to public research brain-drain?," arXiv, last revised February 15, 2021, https://arxiv.org/abs/2102.01648.

114 As a matter of illustration, one co-author (Chris Russell) of a famous and influential paper by Oxford Academics recently moved to Amazon as a senior applied scientist; see Stephen Zorio, "Machine Learning: How a paper by three Oxford academics influenced AWS bias and explainability software," Amazon, April 1, 2021, https://www.amazon.science/latest-news/how-a-paper-by-three-oxford-academics-influenced-aws-bias-and-explainability-software. The paper in question is Sandra Wachter, Brent Mittelstadt, and Chris Russell, "Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI," *Computer Law & Security Review* 41 (July 2021): https://doi.org/https://doi.org/10.1016/j.clsr.2021.105567.

115 Zorio, "Machine Learning," *supra* note 114.

Fourth, considering that AI companies develop and sell AI systems, one needs to be aware of the interest in avoiding any regulatory obstacles that impede the market for AI systems and go against the business interest of companies.

6 Elements and Recommendations for a Potential "Shared Responsibility" Framework between Business and States

This section briefly highlights some elements and principles that should form part of any potential regulatory framework based on "shared responsibility." While it has been shown that some of the broad principles contained in AI principles or Ethical guidelines are reflected in business or soft law frameworks, broad principles are not by themselves ill-suited to achieve the objectives of reducing biases and discrimination. It is rather the non-binding nature coupled with broad principles and a company's freedom to interpret them that makes them less effective. However, broad principles can play a role within a legislative framework when more detailed rules exist that ensure legal certainty and adequate enforcement of the legislative framework. Detailed rules also create an advantageous opportunity for companies to easily ensure compliance, despite having to shoulder a potential increase in regulatory costs due to additional requirements. As a matter of illustration, rather than including the principle of transparency, detailed rules on achieving transparency, such as requirements on documentation, explanations, access to datasets, or the source code of the algorithm, are better suited to support the objectives. The last part of this section will outline some elements and recommendations that should be included in a "Shared Responsibility" framework.

First, technical dialogue between industry representatives and state representatives can prepare the floor for designing a regulatory framework. This will ensure that the specific technologies are understood, and that companies have the possibility to explain their current technical capabilities and the tools they use that address the regulatory goals.

Second, the principles and guidelines developed by businesses need to be assessed and used when designing regulatory frameworks. Elements that are identified as being useful and supporting goals defined in the regulatory framework can be made compulsory in a legal framework.

Third, it is important to move away from abstract concepts and vague narratives, such as transparency, fairness or ethical AI, and move towards concrete and definable concepts like human rights violation, and non-discrimination, gender-based discrimination that can be easily measured and incorporated into a legislative framework.

Fourth, the role of business in each step of the regulatory process needs to be clearly defined, and the risks of private rule setting or regulatory capture need to be addressed from the start. Business input in the form of specialized knowledge and experts is welcome to design and enforce effective legislation, but it should not lead to an imbalance of power that in turn leads to regulation designed and enforced purely in the interest of the companies.

Fifth, a company's involvement in the regulatory process can be counter balanced by the inclusion of independent researchers who mediate between the state and companies in its independent assessment of AI risks, compliance, and through expert advice. Such a role should be exercised either by Open Access processes where algorithmic code or datasets are either publicly available or specifically made available to researchers to test, verify, or audit AI systems in line with regulatory requirements.

Sixth, the basis for enforcing discrimination claims in the algorithmic age is a right to know about an AI generated decision.[116] Without knowledge, no further enquiries in terms of evidence gathering, access to algorithmic design or datasets, evidentiary thresholds are necessary. In addition to a right to know about an AI decision, clearly defined principles should be established in the legislative framework to allow victims of discrimination and/or administrations, courts or independent researchers as experts to have access to algorithms and the underlying (training)datasets. To complement access to evidence, rules on the burden of proof should be designed in a way that takes account of the imbalance of power between tech companies and potential victims of discrimination, both in terms of the access and expertise necessary to evaluate a claim of discrimination. An automatic reversal of the burden of proof will most likely be an adequate tool to facilitate the preparation of a claim for algorithmic discrimination.

Seventh, a definition of algorithmic discrimination would need to be incorporated to supplement existing non-discrimination law frameworks with regard to the specificities and peculiarities of AI systems. In this context, the risks of algorithmic discrimination should be reflected in the regulatory toolbox.

116 See for example Norwegian Consumer Council, "Ghost in the machine," *supra* note 19, at 59, that specifies that "Consumers must have the right to object and to an explanation whenever a generative AI model is used to make decisions that have a significant effect on the consumer." But these and similar calls for such principles, which are fundamental, tend to forget that the right to know is the first step before objecting to an AI decision or even launching a complaint or achieve legal remedies.

Eighth, the legislative framework should be accompanied by a principle that addresses the gender gap in AI and enables more diversity and inclusion in the AI workforce.

Ninth, concrete guidance requirements for AI developers should be incorporated as a binding requirement for companies. Specific guidance to developers should include the main elements of a non-discrimination and diversity perspective, like the issues of bias and discrimination at the design stage of building algorithms.

Tenth, in light of the discussed imbalance and distribution of knowledge and understanding of AI systems, it is advisable to envisage the creation of dedicated knowledge centers on the regulators side or equip regulators with sufficient staff and resources to be able to sufficiently assess and regulate AI systems.[117] Only these kinds of centers can enable regulatory bodies to fulfill their tasks effectively when confronted with industry knowledge and thereby diminish the risk of regulatory capture.

7 Summary and Concluding Remarks

This chapter aimed to show the role of businesses and states in regulating algorithms and preventing algorithmic discrimination. It was argued that while businesses should be involved in the regulatory process, self-regulation and non-binding AI principles are not the preferred option, and should be disregarded in favor of binding legal rules that can be enforced to ensure the protection of human rights and non-discrimination.

It contributed to the debate on regulating algorithms by focusing specifically on the role of businesses, not only with regard to their future obligations imposed by forthcoming legislative frameworks, but also their role in avoiding discriminations from occurring throughout the lifecycle of AI products that they design. Shedding some light and leading to a more nuanced and balanced view on what businesses can and should do in order to safeguard human rights of those affected by the use of algorithms was one goal of this chapter. In this spirit, it argued that regulators around the world should try to involve businesses in the development of rules by being aware of the limits and the private interests of those who dispose of valuable knowledge for

117 The Norwegian Consumer Council is suggesting in its AI report that "[t]ransnational and national technological expert groups should be established to support enforcement agencies in enforcement endeavors," and that "[e]nforcement agencies must have all necessary resources to enforce infringements." *Id.* at 60.

TABLE 5.4 Key elements and recommendations of a "shared responsibility" framework on algorithmic discrimination

	Principles and recommendations of a "shared responsibility" framework
1	Dialogue at technical level between state and business
2	Incorporation of ideas of AI principles of business into regulation
3	Using concrete and definable concepts instead of abstract concepts
4	Defining the role of business throughout the regulatory lifecycle of AI
5	Role of independent AI experts for regulation and implementation
6	Right to know about an AI decision
7	Definition of algorithmic discrimination
8	Accompanying non-legislative measures on Gender AI Gap and Diversity
9	Concrete guidance document for AI developers
10	AI knowledge centers for regulators

regulators – knowledge that is not always easy to decipher in a world where much of the research on AI is influenced or shaped by researchers affiliated to those business who create the algorithms.

The essence of the challenge of how best to address algorithmic discrimination when confronted with the choice between non-binding guidelines and legally binding norms has been highlighted by the UN Special Rapporteur on racial discrimination, "Ethical approaches to governing emerging digital technologies must be pursued in line with international human rights law, and states must ensure that these ethical approaches do not function as a substitute for development and enforcement of existing legally binding obligations."[118] Ethical guidelines and approaches by businesses should also be no substitute for future specific regulatory frameworks as has been argued throughout this contribution.

118 HRC, "Racial discrimination," *supra* note 16, at para. 45.