Article | Received 16 April 2025; Revised 9 June 2025; Accepted 18 August 2025; Published date
https://doi.org/10.55092/let20250007

Weight of gender in artificial intelligence models' implementation in the European Union non-discrimination laws

Angèle Albrengues and Lerong Lu*

The Dickson Poon School of Law, King's College London, UK

* Correspondence author; E-mail: lerong.lu@kcl.ac.uk.

Highlights:

- Critically assess the efficiency of EU non-discrimination laws.
- Highlights the flaws of EU laws to combat gender discrimination in AI.
- Proposes prevention and architecture-based measures to complete the laws.

Abstract: This paper examines the intersection of gender discrimination and artificial intelligence (AI) within the context of European Union (EU) non-discrimination laws. As AI becomes increasingly integrated into our everyday lives, concerns over its potential to perpetuate or exacerbate gender bias have emerged. This work aims to determine whether the current EU legal framework adequately addresses the risks of gender discrimination associated with AI models. Using an interdisciplinary approach, combining law and technology, the study critically assesses the robustness of current EU non-discrimination laws in mitigating AI-induced gender biases. It also explores how these laws could be enhanced to better protect against such discrimination. Methodologically, this paper relies on legal analysis and literature review to identify gaps in the current framework and to propose more effective regulatory and preventive measures. The key findings indicate that, while EU non-discrimination laws provide a foundational framework, they fail to address the unique challenges arising from the implementation of AI models. The implications of these findings suggest that a more comprehensive approach is needed to reduce gender discrimination in digital environments, including legal reforms, technological safeguards, and preventive measures.

Keywords: artificial intelligence; gender discrimination; regulatory framework; European Union; bias; equality by design; EU AI Act; Digital Services Act; GDPR

1. Introduction

The United Nations Office of the High Commissioner for Human Rights has highlighted that women and girls cannot escape sexism, misogyny, and gender-based violence in digital spaces [1]. Despite initial hopes that the internet would be a utopia, online interactions are still plagued by the same inequalities and violence found offline, posing significant risks [2–5]. Artificial intelligence has become a breakthrough technology, with its rapid growth promising improvements in precision, efficiency, and speed across a broad range of applications. However, certain AI models also introduce new risks,

particularly concerning discrimination, which disproportionately affects vulnerable groups like women. Consequently, they often reinforce existing inequalities while creating new ones [6]. For instance, Reuters reported in 2018 that Amazon attempted to develop a CV-screening tool using AI trained on resumes collected over the past decades [7]. Since most of these resumes were from men, the AI system subsequently discriminated against women, leading to the tool's abandonment. Therefore, it becomes increasingly urgent to address and prevent discrimination to protect minorities from unfair treatment in the digital realm.

There is no universal definition of AI, as it often serves as a broad term encompassing various techniques [8]. The latest version of the European Union AI Act, proposed by the European Parliament and the European Council, defines AI as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments"[1]. Similarly, the Organisation for Economic Co-operation and Development (OECD) defines AI models as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments"[2]. A branch of AI is machine learning, which involves discovering relationships "between variables in a dataset, often to make predictions or estimates of some outcome"[3]. AI techniques are used in various applications, including search engines, image and speech recognition systems, and self-driving cars. In this study, the terms AI model, machine learning, and algorithmic decision-making will be used interchangeably for readability purposes.

This paper specifically addresses gender discrimination caused by AI models. Gender discrimination in AI will be defined as when one person is treated less favorably based on sex or gender by an AI model or when an apparently neutral provision, criterion, or practice would put persons of one sex or gender at a disadvantage compared with persons of the other sex or gender, drawing from the Gender Equality Directive definitions of discrimination. Notably, the EU recognizes sex as the foundation for exclusion and bias, implying that an individual's sex is the determining factor for categorizing someone as a woman. However, in all ensuing documents and AI-related reports that will be examined later in this analysis, there is a recurring mention of the word "gender" instead of "sex". This change in terminology could signify a change of viewpoints in the EU, making the definition of what constitutes a man or a woman broader, beyond biological and physiological characteristics. In this analysis, the word "gender" reflects the broader and more current comprehension of this notion.

This paper will examine the robustness of the EU's legal framework in the face of the discrimination risks posed by AI. Are current EU laws sufficient to tackle gender discrimination by AI models? If so, how effective are they? What improvements could be made? These are some of the questions this paper will address, focusing on the impact of gender in AI implementation and the effectiveness of EU non-discrimination laws. Our principal conclusion is that, while the EU's existing legal framework regarding non-discrimination is a positive step towards a more equitable digital environment, limitations are evident. Therefore, a revision of legislation, alongside technological measures and preventive strategies

[1] Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts, [2021] "The EU AI Act", Article 3(1), see [9].
[2] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, (2024), accessed in August 2024 at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449, see [10].
[3] Lehr D.,Ohm P., "Playing with the data: what legal scholars should learn about machine learning", (2017) 51(2) U.C. Davis Law Review, 671, see [11].

is required to fully address the realities of discrimination in AI. In other words, as Xenedis *et al.* suggest, while the EU framework provides "useful yardsticks"[4], there is still room for improvement [13].

This interdisciplinary paper will bridge law and technology and contribute to the ongoing debate on biases within AI and the necessary regulatory framework. Although the EU leads in promoting ethical AI, through initiatives like the EU AI Act, there are significant gaps and weaknesses in its approach [13]. There is an emerging consensus that AI models present significant threats to the legislation on non-discrimination, and, although the topic of gender discrimination and the use of AI models is being increasingly explored in a wide range of sectors, it remains unclear whether the existing EU legal framework is adequately equipped to address these issues. It is therefore crucial to examine this region and determine whether the current laws are strong and flexible enough to meet the complex and ever-changing challenges presented by AI [14]. Therefore, this paper will critically evaluate the resilience of the EU's non-discrimination laws in the face of AI-driven discrimination. When it comes to AI and discrimination, a vast majority of authors agree that AI-powered models can discriminate [15]. However, AI could be used to reduce the probability of discrimination [12]. By leveraging AI to identify biases, increase transparency in decision-making processes and combining it with preventive measures and legal reforms, a more effective framework for tackling gender discrimination in AI could be established. Although the solutions suggested to address the challenges of ethical AI vary widely, we argue that, even if revised, the legislation would need to be complemented by both preventive and architecture-based solutions. Therefore, we emphasize that the implementation of well-crafted feminist policies and a rethink of online spaces' design could be a quicker and more efficient avenue to help reduce gender-discrimination associated with AI models while still recognizing the importance of legal reforms. This strategy will foster diversity and inclusion, an effort that will be assisted by the integration of AI tools.

While this paper aims to provide a comprehensive analysis of gender discrimination in AI within the EU legal framework, it does not extensively compare it with frameworks from other regions, such as the United States or Asia, which may offer alternative approaches to tackling AI-driven discrimination. Furthermore, while the discussion acknowledges the evolving definitions of gender in legal and AI contexts, it does not delve deeply into the intersectionality of gender discrimination with other factors such as race, disability, or socioeconomic status.

The second section of this paper analyzes the non-discrimination laws in the EU, showing that they represent a first step towards a more inclusive AI, while the third section highlights that this framework by itself is insufficient to combat gender discrimination in AI. As a result, the fourth section shows that it is necessary to find new ways to fight against gender discrimination in the implementation of AI models in the EU.

[4] Xenidis R. and Senden L., "EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination" (2020) Ulf Bernitz *et al* (eds), General Principles of EU law and the EU Digital Order, Kluwer Law International, 151–182, see [12].

2. Legal background: a review of EU non-discrimination laws

2.1. Including AI in EU non-discrimination laws

EU non-discrimination laws comprise two key components: primary law, which offers a solid foundation for the development of a more comprehensive and suitable framework, and secondary law, which provides more direct avenues for addressing gender discrimination in AI.

2.1.1. Primary EU non-discrimination laws

The EU's primary non-discrimination laws serve as a robust foundation upon which future AI laws can be built, drawing from the established principles of equality. These primary laws, which are binding agreements among EU Member States, outline the objectives of the EU, dictate the rules for EU institutions, and define the decision-making processes and relationships between the EU and its Member States [16]. Since 2009, these laws have been legally binding for both EU institutions and the authorities of the Member States [17].

Regarding gender equality, the Charter of Fundamental Rights of the European Union explicitly addresses equality in its third chapter. The Charter is an important piece of legislation in the EU. It is a contemporary and all-encompassing tool that safeguards and advances individuals' rights and freedoms, considering societal changes, social progress, and advancements in science and technology. It unifies the freedoms and rights of EU citizens into one legally binding document. Article 21 of the Charter, which focuses on non-discrimination, provides a comprehensive and sector-neutral list of prohibited forms of discrimination, including discrimination based on "sex"[5]. Furthermore, Article 23 emphasizes equality between men and women[6]. These articles were both invoked by the Court of Justice, affirming that "Equality between men and women is a fundamental principle of the European Union" and prohibiting "any discrimination on grounds of sex" requiring "equality between men and women to be ensured in all areas"[7]. The provision for gender equality in Article 23, in addition to its inclusion in Article 21, underscores the historical significance of gender as a basis for status discrimination in EU law [19]. Because of their binding characteristics, these laws both provide a fertile ground for AI legislation to emerge, while putting in place a general framework to fight against discrimination that can be used during the wait for adequate laws tailored to new technologies. As a result, the values enshrined in the Charter serve as guiding principles in the development of AI laws.

2.1.2. Secondary EU non-discrimination laws

Secondary EU law comes from the principles and objectives set by the treaties enacted by the EU. Although these include legislative and non-legislative acts, this analysis will solely focus on legislative acts. The EU can enact three types of binding acts, regulations, directives and decisions. While

[5] Notably, EU laws recognize sex as the foundation for exclusion and bias, implying that an individual's sex is the determining factor for categorizing someone as a woman. However, in all ensuing documents, recent CJEU cases and AI-related reports that will be examined later in this analysis, there is a recurring mention of gender instead of sex. This change in terminology could signify a progression of the viewpoints. In this analysis, we will use the term 'gender' to reflect the more current comprehension of this notion but will use the word "sex" when citing the Charter.
[6] Here, under-represented sex refers to women.
[7] *Association Belge des consommateurs Test-Achats and Others v Conseil des ministres*, [2011], C-236/09 ECR I-77, European Court of Justice, see [18].

regulations are binding in their entirety to all EU Member States without having to be transposed into national law, directives set objectives for EU Member States to achieve a goal, leaving them free to choose how to achieve these objectives before being transposed. Finally, decisions are entirely binding to EU Member States, unless specifying to whom it is addressed. In pursuing a more robust framework to combat gender discrimination in AI, the EU anti-discrimination directives present valuable tools for direct action, even by setting a minimal standard and establishing broad legislation that needs to be transposed into each Member State's domestic law [20]. European non-discrimination laws differentiate between two types of discrimination through the concept of neutrality: direct discrimination, referring to adverse treatment based on a protected attribute such as sexual orientation or gender, and indirect discrimination which describes a situation where an "apparently neutral provision, criterion or practice"[8] disproportionately disadvantages a protected group in comparison with other people [23,24]. This distinction initially emerged from the European Court of Justice jurisprudence, particularly cases such as the *Bilka-Kaufhaus* case, which recognized that facially neutral practices, including a criteria excluding part-time workers from specific benefits, could still result in disproportionate disadvantages for protected groups, most notably women. This distinction was later codified in Directive 97/80/EC on the burden of proof in cases of discrimination based on "sex" in Article 2, to standardize its application across Member States [25]. Since then, the EU has enacted four anti-discrimination directives which adopt the same definition: the Racial Equality Directive, the Gender Equality Directive, the Gender Access Directive, and the Employment Directive [23,24,26,27]. Only two directives focus on gender. Beyond simply mentioning that equality should be ensured, the non-discrimination directives provide a more tailored approach, targeting specific sectors. The Gender Equality Directive mandates equality in the workplace regarding opportunities and treatment of men and women [28]. Similarly, the Gender Access Directive provides a framework for combating discrimination based on "sex" in access to and supply of goods and services [28]. Therefore, they provide a more targeted approach than the Charter, focusing on specific sectors, which is particularly relevant as AI permeates every sector, and can be leveraged to tackle gender discrimination.

Under secondary non-discrimination laws in the EU, gender discrimination is prohibited when a specific action constitutes an infringement of gender equality regulations, such as not hiring a woman due to pregnancy [29,30]. As a result, this legal classification should apply equally to algorithmic discrimination, for example, when businesses employ algorithms for hiring processes [13]. Therefore, EU non-discrimination legislation presents an opportunity to tackle algorithmic bias within the labor market and in the access to and supply of goods and services [12]. The principle of equal pay, now incorporated in Article 157 of the Treaty on the Functioning of the EU, and Article 4 of the Gender Equality Directive, serves as an illustrative case. This principle mandates equal remuneration for women and men engaging in equivalent work or work of equal value. The *Bilka-Kaufhaus* case determined that excluding part-time workers from specific pay or benefits could amount to indirect "sex" discrimination if the exclusion disproportionately impacts women compared to men [31]. This decision brings into question the legality of using algorithmic scoring to determine pay when it adversely affects protected groups, such as gender minorities [12]. Although it is important to remember that this principle is imperfectly applied in non-digital industries, and that it becomes even more complex to apply it within

[8] The Gender Equality Directive, Article 2(1)(b); The Gender Access Directive, Article 2(b), see [21,22].

the digital economy, these directives appear to lay a solid foundation for safeguarding gender minorities' rights against discrimination in AI in specific areas [12].

2.2. Including non-discrimination provisions in technology-focused laws

To effectively investigate the integration of non-discrimination provisions in the new legislation targeting the use of AI we must consider the recent Artificial Intelligence Act together with some older legislation on the use of technology.

2.2.1. The EU AI Act

The EU AI Act presents a unique opportunity to tackle gender discrimination in the sector, using the primary and secondary laws as initial grounds to create a text tailored for AI. The EU AI Act restricts and regulates certain types of behaviors regarding gender equality, such as recruitment processes [32,33]. Although gender is not explicitly mentioned in the core of the Act, its Article 1 states that its purpose is to "improve the functioning of the internal market, promote human-centric and trustworthy AI", and ensure a high level of protection for health, safety, and fundamental rights enshrined in the Charter, which includes "sex" equality. According to its Article 2(7), the Act applies to anyone who makes, uses, imports, or distributes AI systems in the EU, regardless of where they are based. It also applies to AI systems used in the EU, even if they are made elsewhere. The majority of the Act references to non-discrimination and gender can be found in its Recitals, which mention discrimination 22 times, and gender in Recitals 27, 48, 58 and 165. This sets out the objective pursued by the Act and can help judges and law practitioners to interpret and understand the essence of the Act's provisions [34]. Meanwhile, the Act's body prohibits certain AI applications and defines high-risk AI systems in Articles 6, 6(2), and Annex III [32,35]. According to the Act, an AI system is deemed high-risk if it functions as a safety component within a product or if it is a product itself that falls under EU regulations. Certain AI systems are always classified as high-risk unless they do not present a significant threat to individuals' health, safety, or rights. Moreover, if gender discrimination occurs in areas defined in Annex III, such as biometric identification, education, employment, or access to essential services, it could fall under the high-risk category [35,36]. The Act goes even further in its Article 7(b) by adding the possibility to add some high-risk AI to Annex III, when a system poses a "risk of adverse impact on fundamental rights", taking into account "the extent to which potentially harmed or adversely impacted persons are in a vulnerable position in relation to the user of an AI system in particular due to an imbalance of power, knowledge, economic or social circumstances, or age"[9], in its paragraph 2(f). AI systems used for profiling or assessing whether an individual will receive a job or a credit could be considered high risk. As a result, this open list could include gender discrimination more broadly than the initial list set up in Annex III and Article 6, as there is a reference to a vulnerable position due to social circumstances or imbalances in power, characteristics that can exist between different genders.

In cases where gender discrimination falls under the high-risk category, the Act could ensure compliance from the deployers of these systems. Article 8 of the Act mandates compliance with specific requirements outlined in Articles 9 to 14. These applications include, but are not limited to, risk management systems, human oversight, validation and testing of data sets, and compliance with data

[9] The EU AI Act Annex III, see [37].

and data governance principles in terms of training. The obligations of providers and users of high-risk artificial intelligence systems include establishing a quality management system such as strategies for regulatory compliance, design and development procedures, testing and validation processes [38,39]. For example, drawing up the technical documentation for a high-risk artificial intelligence system, such as technical documentation or EU declaration of conformity, keeping the logs automatically generated by high-risk AI systems, and ensuring relevant conformity with the assessment procedure [40,41]. Hence, if a company realizes that its system is not following the rules set by the EU, it should fix the issue, stop using it or recall it, and inform anyone involved in distributing or using the system [42]. The Act also proposes sanctions and the creation of a European AI Office [43–45]. This office is tasked with creating and monitoring guidelines on how to use AI in the EU, to keep information up to date, to describe the data used to train AI, and how to identify risks. Therefore, these provisions could guarantee adequate oversight of AI models and empower relevant authorities to confirm compliance with the Artificial Intelligence Act [13]. While the horizontal nature of the Artificial Intelligence Act and its market-based approach do not specifically address discrimination and gender equality outside of the non-operative part, the Act provides valuable tools for combating gender discrimination in AI [13].

2.2.2. The DSA and the GDPR

In addition to the EU AI Act, other tech-centered regulations may offer valuable tools to address gender discrimination in AI. The DSA applies to intermediary services offered to recipients of the service in the EU. It could be a useful tool, as it mentions gender discrimination. In its explanatory memorandum, the DSA acknowledges the risk of gender discrimination, noting that specific groups may be disadvantaged in their use of online services because of their gender [46]. It highlights that this category may be unfairly impacted by constraints resulting from biases potentially ingrained in user and third party-generated notification systems, as well as perpetuated in automated content moderation tools employed by platforms. The Regulation of the European Parliament and The Council on the DSA of the 19th October 2022 incorporated the right to gender equality and non-discrimination in recitals 3, 52, 81, 83, 134, and Articles 34 and 48 [37]. It seems that the DSA could be used to control AI models once they are put into place. For instance, Article 34 requires providers of very large online platforms and online search engines, including algorithmic systems, to conduct risk assessments when there are any actual or foreseeable negative effects for the exercise of fundamental rights, such as the non-discrimination principle enshrined in Article 21 of the Charter. Article 48 recommends the drawing up of a crisis protocol that safeguards to address any negative effects on the exercise of the fundamental rights of the Charter. If properly enforced, these provisions could establish significant benchmarks for tackling some of the most widespread issues within the digital ecosystem [47].

Moreover, the GDPR includes provisions that could support individual rights in the context of gender discrimination in AI models [7,48]. The GDPR applies to "the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system"[10]. It lists special categories of data, governed by stricter principles. This could, in part, be explained by a wish to prevent unfair discrimination. Although these special categories of personal data do not comprise gender, in

[10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, Article 2(1), see [49].

some circumstances, gender could be considered as such if it was broadly interpreted under the categories "health data", "biometric data" or "genetic data" [50]. Additionally, the GDPR recognizes the freedoms and principles of the Charter and the CJEU recently linked gender and GDPR in C-247/23 Deldits. The court ruled that the GDPR requires the data controller to rectify the gender identity data, when these data are inaccurate, to ensure a high level of protection of fundamental rights and freedoms of natural persons, thereby directly referencing the Charter and confirming that the GDPR is not only about data protection but also about protecting fundamental rights [51]. As a result, certain provisions could help combat gender discrimination by monitoring the implementation of AI models and safeguarding individuals' rights. For instance, Article 22 safeguards individuals from automated decision-making, including fully automated credit scoring and e-recruiting processes, which can have discriminatory outcomes [52,53]. Various other safeguards apply to these data processing activities and could be useful in the context of the development of AI models. This encompasses the general data processing principles outlined in Article 5, the legal grounds for processing described in Article 6, regulations for processing special categories of data in Article 9, transparency and access requirements related to algorithmic decision-making covered by Articles 13 to 15, and the obligation to do data protection impact assessments in specific situations as detailed in Article 35.

As a result, non-discrimination provisions in tech-focused laws, such as the Digital Services Act and the GDPR, and traditional EU non-discrimination laws seem to offer a promising path towards mitigating gender discrimination in AI models and fostering a more equitable digital landscape. However, these frameworks have significant shortcomings both because of their scope and because of the way they are enforced.

3. Critical analysis and challenges

3.1. Conceptual challenges in EU non-discrimination laws

In addressing the conceptual challenges of EU non-discrimination law, it is necessary to consider the shortcomings of this framework's scope to show how these conceptual issues risk normalizing some problematic systems.

3.1.1. Shortcomings of the framework's scope

The EU non-discrimination laws, both when it comes to the traditional and the tech-focused provisions, have a limited scope, preventing the framework from being completely effective.

To begin with, the original EU non-discrimination laws are ill-suited to address algorithmic discrimination due to their paradoxical broad and yet limited scope [54]. AI models often bypass the material and personal scope of anti-discrimination laws, which only cover certain transactions and specific actors. Regarding the material scope, the directives mainly cover employment and access to and supply of goods and services available to the public. While the application for employment does not pose too many issues when it comes to algorithms and even seems to provide some "useful yardsticks"[11], it is more complicated for goods and services. For the directives to apply, algorithms need to be deployed

[11] Xenidis R. and Senden L., "EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination" (2020) Ulf Bernitz *et al* (eds), General Principles of EU law and the EU Digital Order, Kluwer Law International, 151–182, see [12].

for access to and supply of goods and services that are available to the public. However, algorithms are hard to define within these categories. First, the "services" are normally provided against remuneration under EU law [55]. Yet, new types of services have emerged [56]. For instance, some services require customers to pay directly with their data. It is unclear whether consenting to data processing constitutes monetary remuneration under anti-discrimination law. Given the profit-seeking nature of this remuneration, as the data will then be sold by the provider of the service, some scholars argue that anti-discrimination laws should adapt to this development, making these transactions "services" under EU law [54]. Furthermore, the second criterion of availability to the public could also pose some issues. The question would be whether offers to individual persons can be considered as available to the public. Some AI-powered technologies target specific individuals to sell their services. Although the question is debated, these offers should be regarded as publicly available regarding anti-discrimination laws [54]. The key factor should be whether the provider is open to contracting with a broad range of individuals. This perspective highlights the need to concentrate on entities that provide certain types of goods and services to the general public, even if the specific conditions differ from one recipient to another [54]. However, the answers to these questions are not set, preventing the non-discrimination laws from being fully effective.

Regarding the personal scope of EU anti-discrimination legislation, it applies to supply-side actors [54]. However, identifying demand-side actors and intermediaries under this legislation is more challenging. EU anti-discrimination laws do not apply to demand-side actors, such as consumers. The reason for this exemption is to ensure demand-side actors' independence in market decisions, even if they differentiate based on criteria that would otherwise be illegitimate. However, with machine learning tools enabling consumers to delegate purchase decisions to digital personal assistants, this rationale no longer holds in modern markets. As a result, algorithmic consumers should be bound by anti-discrimination laws, as suggested by the CJEU [54]. The increasing use of algorithms disrupts traditional laws to detect, investigate, prevent and correct discrimination [20].

Similarly, the scope of tech-oriented laws appears limited in addressing discrimination. The new EU AI Act for instance, often falls short in effectively combating gender discrimination in AI. Some algorithms used by public or private operators for preparatory purposes are excluded from the Act and not directly accountable for discriminatory actions [57–59]. These algorithms may prompt decisions but do not themselves constitute discriminatory acts. In the event of a discriminatory action on an employee, the discrimination would need to be demonstrated as if the decision had been made by a human being [13]. This coverage gap can leave significant instances of gender discrimination unaddressed. Even the GDPR, while providing valuable protection, has a limited scope in fighting discrimination in AI, as it primarily applies to personal data. Consequently, many AI models fall out of the reach of data protection laws, especially when it comes to decision-making tools. Predictive models, which do not pertain to identifiable individuals, remain unregulated by data protection law [60]. This exclusion creates a loophole where algorithms can operate without accountability, potentially leading to discriminatory outcomes that are difficult to challenge.

3.1.2. Normalization of problematic systems

These conceptual issues extend beyond the scope of EU non-discrimination laws, encompassing problems in the wording and delineation of the concepts inherent to these laws. A preliminary challenge

is the distinction between direct and indirect discrimination. The boundaries between these concepts are blurred because of the complexity of AI models, involving both humans and machines. As a consequence, it has been argued that a "mechanical application of the existing doctrine will fail to address the real sources of bias when discrimination is data-driven"[12]. As a result, some scholars call for a less rigid distinction between the categories. For instance, AG Sharpston indicated in *Bougnaoui* that "it might be objected that the application of the rules laid down by EU law to the latter category[13] is unnecessarily rigid and that some 'blending' of the two categories would be appropriate"[14].

These categories pose inherent problems. Although direct discrimination is expected to be rare when it comes to algorithms, as it involves situations where models are explicitly biased against a protected ground, indirect discrimination will be more prevalent. However, even if the concept of indirect discrimination is helpful, it is weakened by the requirement that the practice must be objectively justified by a legitimate aim [17,54]. Moreover, the means of achieving that aim must be appropriate and necessary [22]. As a result, the strength of the concept of indirect discrimination is compromised by the challenges of this justification [31,63,64]. The "legitimate aim" requirement is relatively straightforward. AI models are often used for legitimate business purposes, such as ranking employees by performance or estimating default risk [54]. In that case, the use of the AI tool must also be appropriate and necessary. The Court will likely consider AI models appropriate as they allow users to make accurate predictions, given their development for precision and granularity beyond human capability [54]. Regarding the necessity requirement, the CJEU mandates that there must be "no other means of achieving the same aim that imposes less of an interference with the right to non-discrimination"[15]. Here, we take the same view as Xenedis *et al.* It seems that the Court of Justice is less likely to accept broad justifications for this last part of the proportionality test [54]. While indirect discrimination can conceptually address a wide range of issues, its practical application is challenging, raising doubts about its effectiveness in systematically addressing algorithmically induced discrimination [12]. This situation opens the possibility for using factors correlated with economically or otherwise favorable traits, even if those factors result in the unfavorable treatment of a protected group. As a result, the necessity requirement would probably be the hardest to prove. Still, the strength of indirect discrimination is reduced, because of the law not being suited for algorithms. As the existing frameworks are not well-equipped to handle the complexities of AI models, the effectiveness of addressing indirect discrimination is significantly diminished. Algorithms can obscure the causes of biased outcomes, making it difficult to demonstrate that there were no less discriminatory alternatives available. This mismatch between the law's requirements and the nature of algorithmic systems weakens the ability to effectively address and remedy indirect discrimination in the digital age.

3.2. Enforcement challenges of EU non-discrimination laws

The enforcement of EU non-discrimination laws is challenging due to the lack of effectiveness of the EU AI Act and the approach taken by the more traditional legal frameworks.

[12] Kim P.T., "Data-Driven Discrimination at Work", (2017) 58(3) William and Mary Law Review, 857, see [61].
[13] The "latter category" refers to direct discrimination.
[14] *Asma Bougnaoui and Association de défense des droits de l'homme (ADDH) v Micropole SA*, [2016] C-188/15 EU:C:2016:553 [65], Court of Justice of the EU, see [62].
[15] European Union Agency for Fundamental Rights and Council of Europe, (2018) "Handbook on European non- discrimination law", accessed in August 2024 at: https://fra.europa.eu/en/publication/2018/handbook-european-non-discrimination-law-2018-edition, see [65].

3.2.1. Lack of effectiveness of the EU AI Act

The risk categorization approach chosen by the EU AI Act has faced criticism. Although the AI Act seeks to establish a thorough legal framework, it faces challenges in addressing the subtleties of algorithmic biases and the intricate nature of discrimination by AI models. Consequently, the Act seems ill-suited to tackle discrimination efficiently. By adopting a list-based approach for high and medium-risk AI systems, the European Economic and Social Committee (EESC) warns that the Act "does not necessarily mitigate the risks of harm to health, safety and fundamental rights for all high-risk AI"[16]. For instance, the AI Act puts in place requirements that are mandatory primarily when it comes to high-risk AI models [67]. This provision includes the possibility of using sensitive data to ensure bias detection and correction [68]. By limiting this provision to high-risk AI systems, it limits the ability to address biases in other AI applications. This restriction may perpetuate discriminatory outcomes in lower-risk systems that still influence societal decisions, creating a risk to normalize biased AI systems [66,69]. As it stands, it might be difficult for the Act to keep pace with rapidly evolving technologies, which could lead to gaps in oversight and control. Similarly, the distinction between categories, particularly concerning banned systems, is not always clear-cut, posing difficulties for both AI developers and regulators. The classification depends on various conceptual categories that merge undesirable effects with specific areas or applications, leading to the inclusion of problematic exceptions [70]. This approach allows developers to determine themselves whether an AI model qualifies as high-risk, potentially weakening the effectiveness of the regulatory framework [71]. This self-identification places critical societal decision-making about human rights protection into the hands of private actors who lack accountability and transparency.

Additionally, the effectiveness of the EU AI Act is limited as it does not go far enough in addressing key issues. First, the Act falls short in promoting gender equality. Originally designed for product safety regulation, the AI Act's framework may not adequately address the unique needs and challenges of regulating AI technologies, particularly regarding gender discrimination [72]. The EU AI Act does not define discrimination or bias in the context of specific challenges for algorithmic discrimination, nor does it stipulate specific principles or measures to combat it [73,74]. Additionally, it uses a gender-neutral language, only using the word "women" twice in its Recitals 56 and 57, therefore overlooking the unique challenges faced by marginalized groups [34]. As a result, the Act does not sufficiently promote equality between men and women. The absence of an equality clause or positive duty for AI models to foster equality is unfortunate. Furthermore, the terms "feminism" or "feminist AI" are never mentioned in the Act. This omission may reflect an intention to avoid politicizing AI, but it overlooks the fact that many AI tools reinforce gender stereotypes, for example by defaulting AI assistants to feminine personas [75,76]. Second, the act mandates transparency, requiring developers and users to elucidate AI decision-making processes and provide evidence to support their methodologies [77]. However, the practical implementation of transparency is challenging. Transparency, even if coupled with data quality, security, human oversight and documentation, is not suitable to counteract discrimination [73]. Ensuring that all stakeholders possess the technical capability to comprehend the code behind the model adds another layer of complexity, especially when

[16] Official Journal of the European Union, "Opinion of the European Economic and Social Committee on Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts" [2021] (COM(2021) 206 final—2021/106 (COD)), OJ C 517, see [66].

transparency is ensured by an explanation given *ex-post* where the harm has already been done and is irreversible [78]. Sometimes, when an explanation is given *ex-post*, it might not even be helpful for individuals [78]. More than that, many AI systems operate as "black boxes", making it challenging to fully understand and explain how they work. When an AI model is referred to as a black box, it means that its internal workings are invisible to the user, and that it is therefore not possible to look at the system's code or logic that produced the output given. When an algorithmic system makes a decision based on analyzing large volumes of data, it is often challenging, if not impossible, to explain the reasoning behind that decision [79–81]. Some scholars even concluded that, when it comes to deep learning, sometimes no one knows exactly how or why it works [82]. As a result, requiring transparency to elucidate AI decision-making processes does not go far enough in protecting gender minorities from discrimination. While the goal of this section is not to review each provision of the EU AI Act, this analysis already shows that it seems to fall short in protecting gender minorities from discrimination occurring in the implementation of AI models.

3.2.2. The traditional EU laws' challenging approach

Ensuring compliance with and enforcement of the EU non-discrimination principle in the context of algorithmic discrimination is challenging, especially due to the mismatch between harmonized EU-level norms and fragmented national enforcement mechanisms. There are two main approaches to enforcement: the individual rights-based approach, in which individuals bring complaints to the courts, and the monitoring and supervisory approach, which involves various public institutions, tools, and mechanisms taking proactive steps to implement and enforce equality. Both approaches highlight significant flaws in the current framework, demonstrating that it is ill-equipped to address algorithmic discrimination effectively. First of all, the individual rights-based approach appears unrealistic for several reasons. The expertise needed to launch a case on non-discrimination law grounds can cause several issues. The first issue concerns the identification of the discrimination itself. When it comes to algorithms, it can be difficult for victims to realize that they were the victims of discrimination due to their opacity [52]. Similarly, establishing the evidence of discrimination could also be hard, in view of algorithmic complexity especially when it comes to direct discrimination, for instance [12,54]. Another issue arises when algorithms are used to mask direct discrimination. This can result in covert direct discrimination, which is more complicated to prove and would likely be considered as indirect discrimination by the Court of Justice of the EU [83]. Further, even when discrimination can be established, accessing justice presents further challenges. Determining liability among the designer, developer, and provider, understanding the applicable laws, and facing potential barriers to accessing information about algorithmic systems due to trade or state secrets, can be difficult. Consequently, many individuals may be deterred from bringing their cases to court [12,80]. Finally, another hurdle is the limited invocability of EU non-discrimination law, depending on whether the defendant is a public or private entity [12]. Despite rulings in cases such as *Mangold, Kücükdeveci, Dansk Industri*, and *Egenberger*, which recognize the horizontal direct effects of non-discrimination, there are still uncertainties about how consumers and users of AI models can enforce their EU rights given existing limitations [84–87]. This protective potential needs confirmation from the Court of Justice [12]. These issues contribute to a lack of clarity and security for claimants, affecting their willingness and ability to launch successful claims related to AI discrimination. When added to other hurdles not specific to EU

non-discrimination laws that claimants face in court, we can question whether it is reasonable to expect people to go to court when it comes to gender discrimination made by AI models [12]. While these problems highlight the barriers at the national level of judicial enforcement, the public supervisory approach reveals similar fragmentation.

Regarding the public supervisory approach, the traditional mechanism to supervise and enforce compliance is the infringement procedure of Article 258 of the Treaty on the Functioning of the European Union, which enables the European Commission to take action against Member States failing to comply with the legislation. However, responsibility for breaches of non-discrimination legislation lies solely with the Member States where EU directives have been transposed into national laws. Consequently, no infringement action can be brought against private entities, even though algorithmic discrimination is likely to occur mainly in the private sector [88]. This situation means that national public bodies must ensure that private actors using AI models are complying with the EU laws. Only these national bodies, not the private entities responsible for the discrimination, can be held accountable by the European Commission for failing to comply. Therefore, the enforcement of the law depends heavily on both national courts and national compliance bodies and agencies, particularly since there is a lack of strong EU supervisory agencies to enforce the law effectively [12]. Given these limitations, the current framework is ill-equipped to address the complexities of AI models' discrimination, leaving significant gaps in oversight and accountability, especially in the private sector. This multi-layered structure creates a system where the EU provides high-level obligations, but no direct mechanism to ensure their uniform, application across Member States.

As a result, the EU non-discrimination framework does not allow for tackling gender discrimination efficiently when it comes to AI models, because of the conceptual challenges of these legal provisions and the enforcement issues it creates. Ultimately, this highlights a key structural tension: while the EU sets horizontal norms to regulate AI-related discrimination, enforcement remains largely vertical, filtered through fragmented national systems that are not fully equipped or harmonized to address the opaque and systemic nature of algorithmic harms. Therefore, it is crucial to identify new strategies to combat gender discrimination in AI models. The next section will explore potential approaches to achieve this goal.

4. The way forward

4.1. Prevention-based measures

Preventive measures can come from the laws but considering alternatives beyond the legal framework is also worthwhile.

4.1.1. Integrating technology and feminism

When revisiting this legal framework, legislators should take both technological advances and gender perspectives into account.

A common view among scholars is that a one-size-fits-all global algorithm law wouldn't be effective [89]. While some argue for creating new sector-specific laws, many believe that adapting existing non-discrimination laws to emerging technologies would be more beneficial [13]. Although sector-specific laws are necessary, they already exist, and adapting them to be flexible, future-proof, and better suited for technological innovations would be more efficient. With improved provisions and

effective enforcement, the EU's legal framework could better combat gender discrimination in AI models' implementation. This analysis does not aim to provide an exhaustive list of measures to enhance the law's effectiveness against gender discrimination, but some suggestions are worth considering. For instance, some organizations, such as the Council of Europe, advocate for revising the presumption of algorithmic bias and the rules on the burden of proof [7]. Revisiting these rules could help restore the effectiveness of non-discrimination laws, especially given the information imbalances between developers and users. This process could allow a shift in the burden of proof onto the defendant when no preventive measures have been taken, without having to establish *prima facie* discrimination in the first place. On a positive note, in April 2025, the Advocate general in C-136/24 Hamoudi/Frontex published an opinion highlighting that the court now tends to shift the burden of proof to the defendant when the claimant has successfully provided *prima facie* evidence to support their case when there is a clear or structural imbalance in access to evidence between the parties [90]. However, even with a well-suited legal framework, the real challenge lies in effective enforcement. By examining gender non-discrimination laws alongside other legal instruments, such as consumer law, competition law, criminal law, and administrative law, we can better combat gender discrimination in AI implementation. For example, consumer law could protect against manipulative algorithmic advertising while competition law could address discriminatory behaviors by monopolistic firms [91–94]. In the public sector, criminal and administrative laws could ensure procedural fairness, which, in this case, refers to the attempts to correct bias in algorithms [95,96]. The GDPR seems to be the most promising tool. It is the first European legislation to recognize algorithmic discrimination and establish rules against it. As a result, it could provide a better-suited monitoring and enforcement toolkit through a preventive framework [12]. This would position it as potentially more effective than traditional anti-discrimination laws, which often rely on individuals to prove harm. Although using the GDPR to combat algorithmic discrimination is "bound to be contentious", scholars have already noted a convergence of anti-discrimination and data protection laws [54]. In particular, the regulation contains two key principles to address algorithmic discrimination: data sanitization, which involves removing sensitive data from datasets used in automated decision-making, and algorithmic transparency, which grants a right to explanation [52,97,98]. In February 2025, the CJEU confirmed in C-203/22, Dun and Bradstreet Austria that individuals subject to automated decision-making have a right to be provided with an explanation that should be relevant, concise, transparent, intelligible and easily accessible, putting the individual in a position to effectively understand it, express their point of view and contest it [99]. These principles allow for earlier identification of potentially biased or discriminatory outcomes during the system design and deployment stages as well as once the algorithm has been deployed. By embedding fairness considerations into the data lifecycle, the GDPR functions as a form of structural governance, requiring organizations to anticipate and address potential discrimination within their system. Combining anti-discrimination laws with the enforcement tools in the GDPR and other legislation could create a powerful mechanism for ensuring fairness in machine-learning contexts. However, the application of these laws to protect people from algorithmic decision-making is still largely unexplored.

Another way to revisit the law is to make it more feminist. As mentioned before, feminism is considered political in the EU. As a result, the concept is not used in non-discrimination laws. However, to reduce gender inequality through AI, certain steps must first be taken in the way legislation is crafted. For instance, Lutz proposes implementing specific gender equality requirements as part of the fight

against gender discrimination [13]. The EU could also adopt a more explicitly feminist approach to AI by increasing the number of references to gender in official documents, laws and reports, and using inclusive language that actively promotes equality and addresses the specific needs of marginalized groups [100]. Additionally, an intersectional approach to gender discrimination is essential, as victims of gender discrimination often face multiple forms of discrimination. Hypothetically, this kind of discrimination could look like a woman customer refused a credit while purchasing goods online, because the combination of her age and gender created an automated rejection from the algorithm. This decision could result from the fact that women around 40 are often divorced and have less economic power [101]. In this example, the model would be reproducing intersectional stereotypes. When gender is combined with other factors, algorithmic decisions can exacerbate social inequality. A woman living in a low-income neighborhood with an unstable job may be charged more for home insurance because her area is deemed high-risk [52]. As someone's financial status is not a protected characteristic, non-discrimination law does not regulate such a practice[17,18]. While empirical evidence remains limited, such crossed-discrimination has happened before, targeting a man, as seen in a decision by the National Non-Discrimination and Equality Tribunal of Finland from 2018. The broader concern is how opaque algorithmic systems risk perpetuating structural inequalities in subtle and unaccountable ways [90,102]. However, it seems more likely for women to end up in such a position, as the percentage of women living in poor households is higher than that of men[19]. As a result, it is paramount to address intersectional discrimination when looking to target gender discrimination in AI models.

Another important issue when wanting to make the law more feminist is defining what constitutes gender and what constitutes discrimination [104]. The scope and limits of non-discrimination directives are not always clear [14]. Historically, the Court of Justice has accepted certain proxies as protected grounds, such as considering discrimination on the grounds of pregnancy as direct "sex" discrimination [31]. However, other cases from the Court's jurisprudence show uncertainty as to which characteristics or combinations fall within protected categories. The degree of overlap required between a proxy and a protected group to constitute discrimination could be clarified. Without a binding provision that defines intersectional discrimination in EU non-discrimination laws, such issues may be overlooked [14]. Therefore, to effectively address gender discrimination in AI, the EU must adopt a more feminist and intersectional approach in its legal frameworks, ensuring that all forms of discrimination are recognized and appropriately regulated.

4.1.2. Beyond legal frameworks

Even if current laws were deemed sufficient, relying solely on legal frameworks would not adequately address gender discrimination in AI models. The law, often grounded in binary ideologies that prioritize male-coded norms, unintentionally reinforces gendered power structures [4]. A critical step towards progress involves challenging and dismantling these ingrained biases, and one way to achieve this is through greater representation of women in the tech sector. The underrepresentation of women in both the professional tech industry and digital spaces influences technology design, making it less responsive to gender and inclusivity [105]. As a result, implementing good practices is a crucial step forward. There is

[17] Here, we assume that the practice does not lead to indirect discrimination based on a protected characteristic.
[18] In data protection law, data about somebody's financial status is not among the "special categories" of data GDPR.
[19] This is called the female poverty rate, see [103].

a need for a paradigm shift in how we design AI models, where inclusionary or exclusionary impacts are systematically assessed [7]. In other words, the deployment of a new AI model should be "purposeful and intentional in its inclusivity" and "must empower communities and present a benefit to all of society"[20]. Good practices could help create a "social justice of AI"[21]. This could take different forms, from conception and development to the testing phase to make sure of the algorithm's fairness [107]. Rather than relying solely on individuals to challenge rights infringements in court, the responsibility should shift to AI developers and deployers, requiring them to meet certain obligations under public scrutiny [108]. These obligations should go beyond simply complying with human rights, and introduce specific standards tailored to societal interests [109]. This could look like developing a label certifying that gender and diversity knowledge is available in a company, mandatory training on gender equality, or, as Virginia Eubanks suggests, developing a new draft of the Hippocratic oath for data scientists system engineers, hackers, and administrative officials [13,110]. Moreover, companies should be enabled to demonstrate compliance with these standards, for instance, through audits of the firm practices and *ex-ante* human rights impact assessments to prevent discriminatory harms [7,107]. For example, the Dutch Platform for the Information Society introduced the Artificial Intelligence Impact Assessment in 2018, highlighting such developments in practice [111]. Similarly, recent tests indicate that the AI model made for art DALLE2, incorporates diversity prompts into vague user questions, such as adding labels like "black" or "female" to a prompt requesting an image of a CEO [112,113]. This method is akin to positive action measures, like quotas. While it can be criticized for not addressing the underlying cause of discrimination, namely, the absence of diversity in training data, when applied on a large scale, this approach has the benefit of promoting more diverse representations, which could help reduce harmful stereotypes over time [7]. However, it is important to avoid "ethics washing" in AI according to Wagner, where firms might view ethics as an easy or soft option [114,115]. While self-regulation has a role, it should not distract legislators from considering new laws or alternative approaches. Encouragingly, recent efforts have been made in this direction within the work of the European Standardisation Organisations, which has been working on guidance for upskilling organisations on AI ethics and social concerns since 2024.

Beyond these good practices, incorporating community-based initiatives is also necessary. Ongoing calls for more stringent or better-enforced laws targeting individual wrongdoers are understandable but ultimately superficial [116]. Prevention strategies must take precedence. Various initiatives, such as the Organisation for Economic Co-operation and Development (OECD) recommendation on AI, the UNI Global Union, and Google AI principles have already been launched. Institutions tasked with monitoring the discriminatory effects of algorithmic decision-making and supporting affected systems should be given priority [117–119]. The Council of Europe advises that these institutions should be endowed with essential legal rights and investigative powers such as access to datasets, adequate resources and funding [7]. Additionally, these institutions should have the capacity to collaborate with users of algorithmic decision-making systems to gather relevant data on the impact of their decisions and to help potential victims seek redress [7]. Existing agencies and organizations such as the European Institute for Gender Equality (EIGE), Equinet, the Fundamental Rights Agency (FRA), the European Data Protection Board, and the recent EU AI Office could map and gather information on AI's adverse effects on society. These

[20] World Economic Forum. How can AI support diversity, equity and inclusion? 2022. Available: https://www.weforum.org/agenda/2022/03/ai-support-diversity-equity-inclusion/ (accessed 14 August 2024), see [106].
[21] Xenidis R., "Tuning EU equality law to algorithmic discrimination: Three pathways to resilience", (2020) 27(6) Maastricht Journal of European and Comparative Law, vol.27, 736–758, see [14].

institutions could lead awareness-raising campaigns, provide holistic AI education, study the impact of AI, develop codes of practice, support individuals' claims, and implement temporary special measures and quotas to promote equality, diversity, and inclusion in the industry [7,104,120,121]. They could also help to bridge the knowledge gap from information asymmetries between AI deployers and impacted individuals, inform the public and pressure the developers to act, with clear established guidelines to follow. Machine-learning researchers, AI developers, and legislators must collaborate with civil society and European organizations[22]. At the same time, social media platofrms and tech companies must provide clear information on how their algorithms work, and allow researches to access these data. Only through this collaboration can the fight against gender discrimination in AI models succeed. The legal field could gain from the technical community's strategies to measure bias, while the technical community needs to consider the adaptable nature of equality in Europe when designing systems. This approach will help to identify emerging forms of discrimination while preserving the contextual judicial interpretation essential to EU non-discrimination laws [20].

4.2. Architecture-based measures

To combat gender discrimination in the implementation of AI models, architecture-based measures are necessary to make wise design choices and learn how to use the technology to our advantage.

4.2.1. Redesigning AI through equality by design

To fight efficiently against gender discrimination in AI models, making wise design choices is paramount. When developing AI models, equality by design should be favored, using a feminist approach. A change in how gender discrimination is tackled is crucial, focusing on a feminist perspective that addresses the structure of online platforms, the business models, and the design strategies used by leading tech companies in our digital world. When thinking about gender and technology, two perspectives coexist: gender in technology and gender of technology [122]. First, gender and technology mutually influence each other. The design of technology not only reflects gender roles and stereotypes but also reinforces them [122]. More specifically, when analyzing some AI models, such as search engine algorithms, and digital platforms, the way gender stereotypes are reproduced and perpetuated through algorithmic design is evident. For example, AI models using facial recognition are more often inaccurate when it comes to the identification of black women, showing the gender and race biases in these technologies [122]. The second concept, gender of technology, refers to the gender associations in the use of technological tools. These associations often assign feminine or masculine characteristics to technology based on societal gender roles [122]. This idea is closely linked to the first concept, as features designed specifically for one gender tend to mirror and reinforce existing gender stereotypes [122]. For instance, devices for household chores are often tailored for women, relying on societal stereotypes, and most voice assistants are designed with female voices, names, and personas, further reinforcing these gendered associations. This demonstrates that the choices made by designers impact the use of specific tools [122]. Developers must take the fact that gender stereotypes are influenced by the interaction between AI tools and users into account when tackling gender-related issues in AI.

[22] Equinet, the umbrella organisation at EU level for national equality bodies, often works on the gendered impact of AI, see [89].

Investing in equality by design to make digital spaces more feminist could be a relevant approach. A proposition, if taking such a line of action, would be to incorporate principles from urban planning to manage the cyberspace. Just as cities do not address every issue by monitoring and arresting every citizen, similar principles can be applied to digital spaces. Movements such as data feminism, feminist data sets, feminist open-source investigations, and feminist principles of the internet offer promising methods to tackle issues such as encryption, anonymity, and gendered harassment [123,124]. These initiatives serve as a form of resistance against the socio-technical power, policies, and politics that shape AI models. They highlight the negative impacts of poorly designed algorithms on gender discrimination and sexism and how societal issues are exacerbated by the values of scale and speed inherent in these technologies [123]. However, there is a significant gap in research that critically examines contemporary surveillance capitalism through a gendered lens. There is a lack of victim-centered policies and decision-making processes that shape technology formats, infrastructure, and business models in line with feminist goals [125]. Addressing this gap is crucial for digital environments that are both incisive and equitable, prioritizing the safety and well-being of gender minorities. When developing AI models, developers should be "purposeful and intentional in their inclusivity" to "empower communities", as this could benefit all of society. This should take precedence over reproducing the bias women are already experiencing in their lives or trying to create a "neutral" space, which would not allow us, as a society, to value the presence of minorities at the margins [106,110,126]. This approach to algorithmic design highlights the importance of incorporating equality and non-discrimination throughout the design and development phases of AI models, which can help establish more robust regulatory and enforcement mechanisms against discrimination [12]. By embracing this concept, decision-makers and legislators can set clear standards for balancing equality and non-discrimination amid technological trade-offs.

4.2.2. Leveraging technology

Using AI as a tool and developing techniques for a trustworthy AI is crucial in combating gender discrimination. Some authors argue that more data or hiding sensitive data is the answer when it comes to fighting gender discrimination in the implementation of AI models [127]. However, this is not enough to efficiently figt against gender discrimination in AI. When data reflects the discrimination that already exists in society, including more data in an AI system will not change the discriminatory output that comes out of it [104]. For example, Criado Perez has revealed that the healthcare industry, by predominantly using male subjects to evaluate drug risks and effectiveness, gathers less and lower-quality data on women's health. This gender data gap results in less reliable predictive systems for diagnosing women [128]. Similarly, removing words such as "women" or "girl" has no effect on the discriminatory outcome of an AI model, as the system can deduce someone's gender by combining other data, as the CV-screening tool evoked in the introduction showed [129]. Instead, Lauren Klein and Catherine D'Ignazio proposed methods for challenging power dynamics in datasets and data projects, identifying biases and using data science to directly confront corporations and governments. They suggested to collect counterdata, analyze data of all kinds with a justice-oriented lens, imagine alternative ways of doing data science, look into the contexts in which those datasets are created and recognize that they are shaped by unequal social relations [104]. Additionally, Wachter *et al.* argue that although AI is responsible for gender discrimination, it could, paradoxically, also be seen as a tool to

fight against it. AI models have the potential to identify biases more effectively and help address structural inequalities [74,130]. The EU views AI as a technology that can enhance human welfare and freedom noting that it "can help to facilitate the achievement of the UN's Sustainable Development Goals, such as promoting gender balance"[23]. Additionally, AI is seen as a means "to achieve a fair society, by helping to increase citizens' health and well-being in ways that foster equality in the distribution of economic, social and political opportunity"[24]. The EU also highlighted that AI was a tool to empower humans and societies [132,133]. These possibilities suggest AI's strategic value in combating gender discrimination, offering both significant risks and promising opportunities in the quest for equality. The issue of discrimination should not be viewed as an inevitable flaw in algorithms, but rather as a trade-off that requires careful consideration. As the statistician George Box famously said, "all models are wrong but some are useful"[25]. The key question is how to define that usefulness and determine which values to prioritize. Often, this trade-off involves balancing ease of access, algorithmic accuracy, and ethical concerns like non-discrimination. In practice, using AI as a tool to fight gender discrimination could look like using AI as an auditor to identify discriminatory patterns and biases and adjusting the model to eliminate them [104,127]. AI image recognition tools could analyze vast amounts of data and assess how gender minorities are represented across different sectors, helping to address discrimination. In content moderation, AI is already used to detect hate speech and it could similarly identify discriminatory language in job advertisements and suggest alternatives [135]. A successful example is Tengai, an AI system implemented in Sweden's public sector to improve recruitment services by reducing biased selection processes [136]. This success demonstrates how AI can be effectively leveraged in the fight against gender discrimination [136].

AI can do more than just function as a tool. It can be guided to become less discriminatory. Increasing rationality and explainability in decision-making, through methods like explainable AI (XAI) or incorporating fairness into models, can help identify correlations, detect imbalanced outcomes, and mitigate bias, making AI safer in addressing gender discrimination [137]. Unlike human decisions, which cannot always be systematically adjusted to test for discrimination, in some cases, AI models can be analyzed, providing better control over how discrimination spreads. In practice, this could look like an obligation to use certain calculation methods to extract the logic of the algorithm, or for developers to use a bias transforming metric that acknowledges historical inequalities and begins from the assumption that certain groups will have a worse starting point than others, instead of using bias preserving metrics that take society as it currently exists as a starting point [74,107]. To take an example from Wachter *et al.*, the latter approach could help to identify talented job applicants who are otherwise undervalued by biased decision criteria that do not accurately reflect merit or competence [74]. Therefore, principles like fairness, explainability, and accountability should be combined with other tools to turn the risks of discrimination into opportunities for greater equality.

Overall, adopting preventive measures while integrating architecture-based solutions seems to be the way forward. A combination of law reforms, preventive measures and the use of technology is necessary to tackle gender discrimination in the implementation of AI models, as shown in Figure 1.

[23] High-Level Expert Group, "Ethics guidelines for trustworthy AI", (2019), accessed in August 2024 at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, see [131].
[24] ibid.
[25] Box G., "Robustness in the strategy of scientific model building" (1979) Robert Launer and Graham Wilkinson editios, Robustness in Statistics, Academic Press, 201–236, see [134].

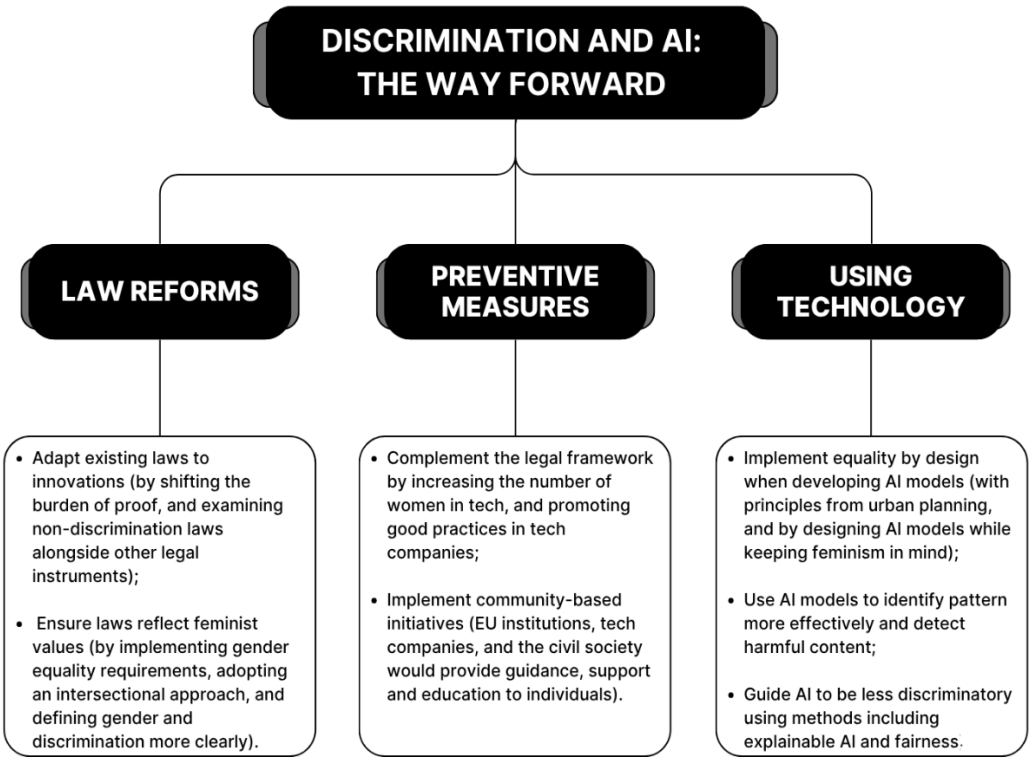

Figure 1. Visual representation of Section 4.

5. Conclusion

This paper has sought to address the critical issue of gender discrimination perpetuated by AI models within the context of the EU non-discrimination legal framework. The research has demonstrated that EU non-discrimination laws provide a solid foundation to fight against gender discrimination, thanks to both primary and secondary laws. Moreover, this framework, when completed with tech-focused legal provisions such as the DSA the GDPR, or the EU AI Act is even more promising. However, these laws are not sufficiently robust to address the unique challenges posed by AI-driven technologies. The interdisciplinary approach employed, combining legal analysis and insights from technology, has revealed significant gaps in the current framework that could allow gender biases to persist and even exacerbate. With both conceptual and enforcement issues, these laws are not adapted to the current reality of the implementation of AI models. This conclusion highlights the need for both preventive and architecture-based measures. Nonetheless, this study has certain limitations. It primarily focuses on the EU framework, without extensive comparisons to non-EU regulations and does not deeply explore the intersectionality of gender bias with other forms of discrimination.

The analysis points to several key policy recommendations when it comes to prevention. Firstly, there is a pressing need for legal reforms that specifically target AI-induced gender discrimination. These reforms should be tailored to the unique challenges of gender minorities coupled with the one of AI models. For instance, they could provide a better definition of gender and discrimination, and adopt an intersectional approach. Secondly, beyond the law, good practices should be developed in the technological sector. With practices such as auditing a firm practice, training on gender equality, and

creating a label certifying that gender and diversity knowledge is available in the company, gender discrimination could be tackled more efficiently. Finally, a collaboration between legislators, developers of AI models, the civil society, and EU organizations is essential in this fight. Providing some institutions such as agencies or the recent AI Office with the necessary funding, but also with legal and investigative powers, could allow them to support individual claims or lead awareness campaigns. However, it is necessary to go a step further than prevention. Architecture-based measures proved to be useful. Making thoughtful design choices to develop equality by design can make digital spaces more feminist. Moreover, using AI as a tool to detect discrimination, to act as an auditor, or leveraging technology by developing techniques for a safer AI, such as explainable AI or fairness, can also be efficient in this combat against gender discrimination in the implementation of AI models.

The objectives of this research were achieved by critically evaluating the existing laws, identifying their limitations, and proposing enhancements that could lead to a more equitable digital environment.

Looking ahead, the rapid evolution of AI technology necessitates ongoing vigilance and adaptation of legal and regulatory frameworks. The EU must continue to lead in promoting ethical AI practices, ensuring that the digital transformation benefits all members of society without reinforcing existing inequalities. By implementing the recommendations outlined in this paper, the EU can pave the way for a more just and inclusive digital future, where AI serves as a force for equality rather than discrimination.

This paper also invites us to reflect deeply on the discrimination that persists in our everyday lives. While algorithms possess the capacity to discriminate, they often mirror the societal gender divide that already exists. Yet, a reciprocal relationship exists: a meaningful shift towards gender equality in the real world will inevitably be echoed in the data we produce, thereby reducing biases in algorithmic outcomes. At the same time, through the adaptation of laws, the cultivation of ethical practices, and the deliberate design of technologies, we can harness these tools to challenge and diminish gender inequality. It is through collaboration and a commitment to learning that we may ultimately bridge the divide, both in the digital realm and in our lived experiences.

Authors' contribution

Conceptualization, Angèle Albrengues and Lerong Lu; methodology, Angèle Albrengues and Lerong Lu; software, Angèle Albrengues; validation, Lerong Lu; formal analysis, Angèle Albrengues; investigation, Angèle Albrengues; resources, Angèle Albrengues; writing—original draft preparation, Angèle Albrengues; writing—review and editing, Lerong Lu; supervision, Lerong Lu. All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The author declares no conflict of interest.

References

[1] United Nations Office of the High Commissioner for Human Rights. UN experts urge States and companies to address online gender-based abuse but warn against censorship. 2017. Available: https://www.ohchr.org/en/press-releases/2017/03/un-experts-urge-states-and-companies-address-online-gender-based-abuse-warn (accessed on 8 June 2025).

[2] Barlow JP. A declaration of the independence of cyberspace. 1996. Available: https://www.eff.org/cyberspace-independence (accessed on 8 June 2025).
[3] Plan International survey. State of the world's girls 2020: free to be online? 2020. Available: https://plan-international.org/publications/free-to-be-online/ (accessed on 14 August 2024).
[4] Arimatsu L. Silencing women in the digital age. *Cambridge Int. Law J.* 2019, 8(2):187–217.
[5] Dunn S. Technology-facilitated gender-based violence: an overview. 2020. Available: https://digitalcommons.schulichlaw.dal.ca/scholarly_works/773/ (accessed on 14 August 2024).
[6] Gerards J, Xenidis R. Algorithmic discrimination in Europe: challenges and opportunities for EU gender equality and non-discrimination law. 2021. Available: https://www.pure.ed.ac.uk/ws/portalfiles/portal/235501180/EELN_report_algorithmic_discrimination.en.pdf (accessed on 5 August 2024).
[7] Bartoletti I, Xenidis R. Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination. 2023. Available: https://rm.coe.int/study-on-the-impact-of-artificial-intelligence-systems-their-potential/1680ac99e3 (accessed on 29 August 2024).
[8] Madiega T. Artificial Intelligence Act. 2023. Available: https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf (accessed on 13 August 2024).
[9] European Union. Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. 2021. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206&utm (accessed on 13 August 2024).
[10] OECD Legal Instruments. Recommendation of the Council on artificial intelligence. 2024. Available: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed on 8 June 2025).
[11] Lehr D, Ohm P. Playing with the data: what legal scholars should learn about machine learning. *UCDL Rev.* 2017, 51:653.
[12] Xenidis R, Senden L. EU non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination. In *General Principles of EU law and the EU Digital Order*, 1st ed. The Hague: Kluwer Law International, 2020. pp. 151–182.
[13] Lütz F. Gender equality and artificial intelligence in Europe: addressing direct and indirect impacts of algorithms on gender-based discrimination. 2022. Available: https://rdcu.be/dRWYe (accessed on 14 August 2024).
[14] Xenidis R. Tuning EU equality law to algorithmic discrimination: three pathways to resilience. *Maastricht J. Eur. Comp. Law* 2020, 27(6):736–758.
[15] Nadeem A, Abedin B, Marjanovic O. Gender bias in AI: a review of contributing factors and mitigating strategies. 2020. Available: https://research-management.mq.edu.au/ws/portalfiles/portal/225086042/224368743.pdf (accessed on 12 August 2024)
[16] European Commission. Types of EU law. Available: https://commission.europa.eu/law/law-making-process/types-eu-law_en#:~:text=The%20body%20of%20law%20that,%2C%20decisions%2C%20recommendations%20and%20opinions (accessed on 1 July 2024).
[17] Official Journal of the European Union. Consolidated version of the treaty on European Union, C326/12, Article 6(1). 2012. Available: https://eur-lex.europa.eu/resource.html?uri=cellar:2bf140bf-a3f8-4ab2-b506-fd71826e6da6.0023.02/DOC_1&format=PDF (accessed on 1 July 2024).

[18] European Union. Association Belge des consommateurs Test-Achats and Others v Conseil des ministres [2011] C-236/09 ECR I-77. 2011. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62009CJ0236 (accessed on 12 August 2024).

[19] Eklund H, Kilpatrick C. Article 21 EU Charter of Fundamental Rights. 2020. Available: https://cadmus.eui.eu/server/api/core/bitstreams/38b35a5a-b514-5a87-a7e3-73c386a4c71f/content (accessed on 8 August 2024).

[20] Wachter S, Mittelstadt B, Russell C. Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI. *Comput. Law Secur. Rev.* 2021, 42:105567.

[21] European Union. 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, Article 2(b). 2004. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32004L0113 (accessed on 1 August 2024).

[22] European Union. 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), Article 2(1)(b). 2006. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32006L0054 (accessed on 29 August 2024).

[23] European Union. Council Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation. 2000. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32000L0078 (accessed on 12 August 2024).

[24] European Union. Directive 2006/54/EC, European Parliament and the Council, implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast). 2006. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32006L0054 (accessed on 12 August 2024).

[25] Wengdhal A. Indirect discrimination and the European Cout of Justice. A comparative analysis of European Court of Justice case-law relating to discrimination on the grounds of, respectively, sex and nationality. 2001. Available: https://lup.lub.lu.se/search/files/5403361/624694.pdf (accessed on 9 May 2025).

[26] European Union. Council Directive 2004/113/EC implementing the principle of equal treatment between men and women in the access to and supply of goods and services. 2004. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32004L0113 (accessed on 12 August 2024).

[27] European Union. Council Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethic origin. 2000. Available:https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32000L0043 (accessed on 2 July 2024).

[28] European Union. 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), Article 1. 2006. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32006L0054 (accessed on 2 July 2024).

[29] European Union. 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), Article 2(1)(a). 2006. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32006L0054 (accessed on 2 July 2024).

[30] European Union. 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), Article 14(1)(a). 2006. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32006L0054 (accessed on 2 July 2024).

[31] European Union. Bilka–Kaufhaus GmbH v Karin Weber von Hartz, [1986] C-170/84 EU:C:1986:204. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A61984CJ0170 (accessed on 4 July 2024).

[32] European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (EU AI Act), Article 6(2). 2024. Available: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (accessed on 2 July 2024).

[33] EU Artificial Intelligence Act. Annex III: high-risk AI systems referred to in Article 6(2), Annex III Nr.4. 2024. Available: https://artificialintelligenceact.eu/annex/3/ (accessed on 2 July 2024).

[34] Karagianni A. The EU Artificial Intelligence Act though a gender lense. 2025. Available: https://library.fes.de/pdf-files/bueros/bruessel/21887-20250304.pdf (accessed on 3 April 2025).

[35] EU Artificial Intelligence Act. Annex III: high-risk ai systems referred to in Article 6(2), Annex III. 2024. Available: https://artificialintelligenceact.eu/annex/3/ (accessed on 2 July 2024).

[36] European Commission. Artificial intelligence—questions and answers. 2023. Available: https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683 (accessed on 10 June 2024).

[37] European Union. 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). 2022. Available: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng (accessed on 2 July 2024).

[38] EU Artificial Intelligence Act. Article 16: obligations of providers of high-risk AI systems. 2024. Available: https://artificialintelligenceact.eu/article/16/ (accessed on 2 July 2024).

[39] EU Artificial Intelligence Act. Article 17: quality management system. 2024. Available: https://artificialintelligenceact.eu/article/17/ (accessed on 2 July 2024).

[40] EU Artificial Intelligence Act. Article 18: documentation keeping. 2024. Available: https://artificialintelligenceact.eu/article/18/ (accessed on 2 July 2024).

[41] EU Artificial Intelligence Act. Article 19: automatically generated logs. 2024. Available: https://artificialintelligenceact.eu/article/19/ (accessed on 2 July 2024).

[42] EU Artificial Intelligence Act. Article 20: corrective actions and duty of information. 2024. Available: https://artificialintelligenceact.eu/article/20/ (accessed on 2 July 2024).

[43] EU Artificial Intelligence Act. Article 71: EU database for high-risk AI systems listed in Annex III. 2024. Available: https://artificialintelligenceact.eu/article/71/ (accessed on 2 July 2024).

[44] EU Artificial Intelligence Act. Article 56: codes of practice. 2024. Available: https://artificialintelligenceact.eu/article/56/ (accessed on 2 July 2024).

[45] EU Artificial Intelligence Act. Article 59: further processing of personal data for developing certain AI systems in the public interest in the AI regulatory sandbox. 2024. Available: https://artificialintelligenceact.eu/article/59/ (accessed on 2 July 2024).

[46] European Union. Proposal for a regulation of the European Parliament and of the Council on a single market for digital services (Digital Services Act) and amending directive 2000/31/EC. 2020. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020PC0825 (accessed on 4 July 2024).

[47] Allen A. An intersectional lens on online gender based violence and the Digital Services Act. 2022. Available: https://verfassungsblog.de/dsa-intersectional/ (accessed on 12 August 2024).

[48] European Union. Regulation EU 2016/769 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). 2016. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng (accessed on 2 July 2024).

[49] European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Article 2(1). 2016. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679 (accessed on 2 July 2024).

[50] Van Bekkum M, Zuiderveen Borgesius F. Using sensitive data to prevent discrimination by artificial intelligence: does the GDPR need a new exception? *Comput. Law Secur. Rev.* 2023, 48:105770.

[51] InfoCuria Case-law. Deldits [2025] C-247/23 EU:C:2025:172. 2025. Available: https://curia.europa.eu/juris/document/document.jsf?docid=296550&doclang=en (accessed on 23 March 2025).

[52] Zuiderveen Borgesius FJ. Strengthening legal protection against discrimination by algorithms and artificial intelligence. *Int. J. Human Rights* 2020, 24(10):1572–1593.

[53] Official Journal of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), Article 22. 2016. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A32016R0679 (accessed on 17 May 2024).

[54] Hacker P. Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Rev*. 2018, 55(4):1143–1185.

[55] European Union. Consolidated version of the treaty on the functioning of the European Union, Article 57. 2016. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A12016E057 (accessed on 21 May 2024).

[56] Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science* 2017, 356(6334):183–186.

[57] European Union. Nadine Paquay v Société d'architectes Hoet + Minne [2007] C-460/06 SPRLECLI:EU:C:2007:601. 2007. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62006CJ0460 (accessed on 2 August 2024).

[58] Official Journal of the European Union. Directive 2019/1158 of the European Parliament and of the Council on work-life balance for parents and carers and repealing Council Directive 2010/18/EU. 2019. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A32019L1158 (accessed on 2 August 2024).

[59] Official Journal of the European Union. Directive 2019/1158 of the European Parliament and of the Council on work-life balance for parents and carers and repealing Council Directive 2010/18/EU, Article 12. 2019. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A32019L1158 (accessed on 2 August 2024).

[60] Wachter S, Mittelstadt B. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. *Colum. Bus. L. Rev.* 2018, 2:443–493.

[61] Kim PT. Data-driven discrimination at work. *William Mary L. Rev.* 2017, 58(3):857.

[62] European Union. Asma Bougnaoui and association de défense des droits de l'homme (ADDH) v Micropole SA. 2016. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62015CJ0188 (accessed on 10 August 2024).

[63] European Union. Dr. Pamela Mary Enderby v Frenchay Health Authority and Secretary of State for Health. 1993. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A61992CJ0127 (accessed on 10 August 2024).

[64] European Union. J.P. Jenkins v Kingsgate (Clothing Productions) Ltd. [1981] C-96/80 EU:C:1981:80. 1981. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A61980CJ0096 (accessed on 21 August 2024).

[65] European Union Agency for Fundamental Rights and Council of Europe. Handbook on European non-discrimination law. 2018. Available: https://fra.europa.eu/en/publication/2018/handbook-european-non-discrimination-law-2018-edition (accessed on 16 August 2024).

[66] Official Journal of the European Union. Opinion of the European Economic and Social Committee on proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. 2021. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AC%3A2021%3A517%3AFULL (accessed on 25 May 2024).

[67] EU Artificial Intelligence Act. Article 8: compliance with the requirements. 2021. Available: https://artificialintelligenceact.eu/article/8/ (accessed on 12 May 2024).

[68] European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Article 10(5). 2024. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689 (accessed on 14 June 2024).

[69] Bosoer L, Gamito MC, Rubio-Marin R. Non-discrimination and the AI Act. In *Law and Digitalization*, 1st ed. Arazandi: Cadaei, 2023. p. 19.

[70] EU Artificial Intelligence Act. Article 6: classification rules for high-risk AI systems. 2021. Available: https://artificialintelligenceact.eu/article/6/ (accessed on 20 June 2024).

[71] EU Artificial Intelligence Act. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). 2024. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689 (accessed on 4 August 2024).

[72] Solano JL, Siddharth de Souza AM, Taylor L. Governing data and artificial intelligence for all: models for sustainable and just data governance. 2022. Available: https://research.tilburguniversity.edu/en/publications/governing-data-and-artificial-intelligence-for-all-models-for-sus (accessed on 14 August 2024).

[73] Berend B. The AI Act Proposal: towards the next transparency fallacy? Why AI regulation should be based on principles based on how algorithmic discrimination works. In *Künstliche Intelligenz–Wie gelingt eine vertrauenswürdige Verwendung in Deutschland und Europa*, 1st ed. Tübingen: Mohr Siebeck, 2022.

[74] Wachter S, Mittelstadt B, Russell C. Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law. *West Virginia Law Rev.* 2020, 123(3):735.

[75] European Institute for Gender Equality. Feminism(s). Available: https://eige.europa.eu/publications-resources/thesaurus/terms/1058?language_content_entity=en#:~:text=Description,their%20gendered%20body%2C%20i.e.%20sex (accessed on 15 July 2024).

[76] Chin-Rothmann C, Robison M. How AI bots and voice assistants reinforce gender bias. 2020. Available: https://www.brookings.edu/articles/how-ai-bots-and-voice-assistants-reinforce-gender-bias/ (accessed on 12 May 2024).

[77] EU Artificial Intelligence Act. Article 13: transparency and provision of information to deployers. 2024. Available: https://artificialintelligenceact.eu/article/13/ (accessed on 12 August 2024).

[78] Edwards L, Veale M. Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. 2017. Available: https://scholarship.law.duke.edu/dltr/vol16/iss1/2/ (accessed on 15 August 2024).

[79] Ananny M, Crawford K. Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *New Media Soc.* 2018, 20(3):973–989.

[80] Burrell J. How the machine thinks : understanding opacity in machine learning algorithms. *Big Data Soc.* 2016, 3(1):2053951715622512.

[81] Hildebrandt M. *Smart technologies and the end(s) of Law*: *novel entanglements of law and technology*, 1st ed. Cheltenham: Edward Elgar Publishing, 2015.

[82] Heaven WD. Large language models can do jaw-dropping things. But nobody knows exactly why. 2024. Available: https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why/ (accessed on 3 August 2024).

[83] Arnadóttir OM. *Equality and non-discrimination under the European Convention on Human Rights*, 1st ed. The Hague: Kluwer Law International, 2003.

[84] InfoCuria Case-law. Werner Mangold v Rüdiger Helm [2005] C-144/04 EU:C:2005:709. 2005. Available: https://curia.europa.eu/juris/document/document.jsf?text=&docid=56394&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=105 (accessed on 13 August 2024).

[85] European Union. Seda Kücükdeveci v Swedex GmbH & Co. KG [2010] C-555/07 EU:C:2010:21. 2010. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62007CJ0555 (accessed on 13 August 2024).

[86] European Union. Dansk Industri (DI), acting on behalf of Ajos A/S v Estate of Karsten Eigil Rasmussen [2016] C-441/14 EU:C:2016:278. 2016. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62014CJ0441 (accessed on 12 May 2024).

[87] European Union. Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung e.V. [2018] C-414/16 EU:C:2018:257. 2018. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62016CJ0414 (accessed on 13 June 2024).

[88] Olsen HP, Slosser JL, Hildebrandt TT, Wiesener C. What's in the box? The legal requirement of explainability in computationally aided decision-making in public administration. 2019. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3402974 (accessed on 26 August 2024).

[89] Smuha NA. Beyond the individual: governing AI's societal harm. 2021, pp. 1–30. Available: https://policyreview.info/articles/analysis/beyond-individual-governing-ais-societal-harm (accessed on 12 Augsut 2024).

[90] Court of Justice of the European Union. Advocate General's Opinion in Case C-136/24 P Hamoudi/Frontex. Action for damages against Frontex: AG Norkus analyses the allocation of the burden of proof in relation to the existence of damage in collective expulsion cases. 2025. Available: https://curia.europa.eu/jcms/upload/docs/application/pdf/2025-04/cp250048en.pdf (accessed on 13 April 2025).

[91] Graef I. *EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility*, 1st ed. Amsterdam: Kluwer Law International, 2016.

[92] Jabłonowska A, Kuziemski M, Nowak AM, Micklitz HW, Palka P, *et al*. Consumer law and artificial intelligence: challenges to the EU consumer law and policy stemming from the business -use of artificial intelligence: final report of the ARTSY project. 2018. Available: https://cadmus.eui.eu/entities/publication/19a962a5-6843-50dd-bb44-50f7f2cdf447 (accessed on 12 May 2024).

[93] Ezrachi A, Stucke ME. *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*, 1st ed. Cambridge: Harvard University Press, 2016.

[94] Valcke P, Graef I, Clifford D. iFairness—constructing fairness in IT (and other areas of) law through intra and interdisciplinarity. *Comput. Law Secur. Rev.* 2018, 34(4):707–714.

[95] Oswald M. Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philos. Trans. R. Soc. A* 2018, 376(2128):20170359.

[96] Cobbe J. Administrative law and the machines of government: judicial review of automated public-sector decision-making. *Legal Stud.* 2019, 39(4):636–655.

[97] Intersoft Consulting. Article 9 GDPR: processing of special categories of personal data. 2016. Available: https://gdpr-info.eu/art-9-gdpr/ (accessed on 12 August 2024).

[98] Zuiderveen Borgesius FJ. Strengthening legal protection against discrimination by algorithms and artificial intelligence. *Int. J. Hum. Rights* 2020, 24(10):1572–1593.

[99] InfoCuria Case-law. Dun and Bradstreet Australia [2025] C-203/22 EU:C:2025:117. 2025. Available: https://curia.europa.eu/juris/document/document.jsf?cid=272718&docid=295841&doclang=en (accessed on 2 March 2025).

[100] Guevara-Gómez A, de Zárate-Alcarazo LO, Ignacio Criado J. Feminist perspectives to artificial intelligence: comparing the policy frames of the European Union and Spain. *Inf. Polity* 2021, 26(2):173–192.

[101] Wulf J. *Automated Decision-Making Systems and Discrimination: Understanding causes, recognizing cases, supporting those affected*, 1st ed. Berlin: AlgorithmWatch, 2022.

[102] Yhdenvertaisuus-ja tasa-arvolautakunta/Täysistunto (ään.). Tapausseloste YVTltk 2xx 2017: Luottokelpoisuuden arviointimenettely. 2017. Available: https://www.yvtltk.fi/material/collections/20230609132927/Hs0XG8IKF/Tapausseloste_YVTltk_2xx_2017_luottokelpoisuuden_arviointimenettely.pdf (accessed on 27 May 2025).

[103] Munoz Boudet AM, Buitrago P, Leroy de la Briere B, Newhouse D, Rubioano Matulevich E, *et al.* Gender differences in poverty and household composition through the life-cycle: a global perspective. 2018. Available: http://documents.worldbank.org/curated/en/135731520343670750/Gender-differences-in-poverty-and-household-composition-through-the-life-cycle-a-global-perspective (accessed on 14 August 2024).

[104] Klein L, D'Ignazio C. Data Feminism for AI. 2024. Available: https://dl.acm.org/doi/pdf/10.1145/3630106.3658543 (accessed on 12 June 2025).

[105] UN Women. Accelerating efforts to tackle online and technology-facilitated violence against women and girls. 2022. Available: https://www.unwomen.org/en/digital-library/publications/2022/10/accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls (accessed on 8 August 2024).

[106] World Economic Forum. How can AI support diversity, equity and inclusion? 2022. Available: https://www.weforum.org/agenda/2022/03/ai-support-diversity-equity-inclusion/ (accessed 14 August 2024).

[107] Jean A. *Les algorithmes font-ils la loi* , 1st ed. Paris: Éditions de l'Observatoire, 2021.

[108] Yeung K. Responsibility and AI—a study of the implications of advanced digital echnoogies (including AI systems) for the concept of responsibility within a human rights framework. 2019. Available: https://rm.coe.int/responsability-and-ai-en/168097d9c5 (accessed on 8 August 2024).

[109] Van der Sloot B, van Schendel S. Procedural law for the data-driven society. *Inf. Commun. Technol. Law* 2021, 30(3):304–332.

[110] Eubanks V. *Automating inequality: how high-tech tools profile; police and punish the poor*, 1st ed. New York: St. Martin's Press, 2018.

[111] ICT Institute. The artificial intelligence impact assessment. 2018. Available: https://ictinstitute.nl/the-artificial-intelligence-impact-assessment/ (accessed on 8 August 2024).

[112] Sparkes M. AI art tool DALL-E 2 adds 'black' or 'female' to some image prompts. 2022. Available: https://www.newscientist.com/article/2329690-ai-art-tool-dall-e-2-adds-black-or-female-to-some-image-prompts/ (accessed on 12 August 2024).

[113] OpenAI. Reducing bias and improving safety in DALL·E 2. 2022. Available: https://openai.com/blog/reducing-bias-and-improving-safety-in-dall-e-2/ (accessed on 25 August 2024).

[114] Greene D, Hoffman AL, Stark L. Better, nicer, clearer, fairer: a critical assessment of the movement for ethical artificial intelligence and machine learning. 2019. Available: https://aisel.aisnet.org/hicss-52/dsm/critical_and_ethical_studies/2/ (accessed on 23 August 2024).

[115] Wagner B. Ethics as an escape from regulation: from ethics-washing to ethics-shopping? 2018, pp. 84–88. Available: https://mediarep.org/server/api/core/bitstreams/33079b45-5f2d-4223-9c48-bdd42955f89b/content (accessed on 29 July 2024).

[116] Jane EA. Bad actors or bad architecture? Rethinking gendered violence online. In *The Routledge Companion to Gender, Media and Violence*, 1st ed. Abingdon: Taylor & Francis Group, 2023. pp. 358–368.

[117] OECD Legal Instruments. Recommendation of the Council on artificial intelligence. 2024. Available: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed on 4 August 2024).

[118] The Future World of Work. 10 principles for ethical AI. Available: http://www.thefutureworldofwork.org/opinions/10-principles-for-ethical-ai/ (accessed on 10 August 2024).

[119] Google AI. Our AI Principles. 2024. Available: https://ai.google/principles (accessed on 10 August 2024).

[120] FRA European Union agency for fundamental rights. #BigData: discrimination in data-supported decision-making. 2018. Available: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf (accessed on 10 August 2024).

[121] EU Artificial Intelligence Act. Article 56: codes of practice. 2025. Available: https://artificialintelligenceact.eu/article/56/?utm (accessed on 10 August 2025).

[122] Faulkner W. The technology question in feminism: a view from feminist technology studies. *Womens Stud. Int. Forum* 2001, 24(1):79–95.

[123] Ganesh MJ, Moss E. Resistance and refusal to algorithmic harms: varieties of 'knowledge projects'. *Media Int. Aust.* 2022, 183(1):90–106.

[124] Massachi S. How to save our social media by treating it like a city. 2021. Available: https://social.cs.washington.edu/cse481social/readings/W9S1/how-to-save-social-media-treat-it-like-a-city-SaharMassachiarchive.pdf (accessed on 12 June 2024).

[125] UNFPA. Preventing Technology-Facilitated Gender-Based Violence (TF GBV). Available: https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/GDC-Submission_UNFPA.pdf (accessed on 14 August 2024).

[126] Pradhan A, Erete S, Chopra S, Upadhyay P, Sule O, *et al.* "No, not that voice again!": engaging older adults in design of anthropomorphic voice asssistants. 2025. Available: https://dl.acm.org/doi/pdf/10.1145/3711039 (accessed on 25 June 2025).

[127] Zou J, Schiebinger L. AI can be sexist and racist—it's time to make it fair. *Nature*, 2018, 559(7714):324–326.

[128] Perez C. *Invisible women: exposing data bias in a world designed for men*. New York: Abrams Press, 2020.

[129] Dastin J. Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of Data and Analytics*, 1st ed. Milton Park: Taylor & Francis Group, 2022.

[130] Margetts H. Rethinking AI for good governance. *Daedalus* 2022, 151(2):360–371.

[131] European Comission. High-level expert group on artificial intelligence. 2019. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed on 13 August 2024).

[132] European Commission. Artificial Intelligence for Europe. 2018. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri= CELEX:52018DC0237&from=EN (accessed on 21 August 2024).

[133] European Commission. Policy and investment recommendations for trustworthy artificial intelligence. 2019. Available: https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence (accessed on 22 August 2024).

[134] Box G. Robustness in the strategy of scientific model building. In *Robustness in Statistics*, 1st ed. London & New York: Academic Press, 1979. pp. 201–236.

[135] European Commission against Racism and Intolerance (ECRI). ECRI general policy recommendation no. 15 on combating hate speech (adopted on 8 December 2015). 2016. Available: https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01 (accessed on 14 August 2024).

[136] Misuraca G, Van Noordt C. AI watch-artificial intelligence in public services: overview of the use and impact of AI in public services in the EU. 2020. Available: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://publications.jrc.ec.europa.eu/repository/bitstream/JRC120399/jrc120399_misuraca-ai-watch_public-services_30062020_def.pdf&ved=2ahUKEwiRj5rMhvCGAxVyWkEAHQaEC_0QFnoECA8QAQ&usg=AOvVaw0Lw25J2TWO43Beo3DwHrCH (accessed on 12 August 2024).

[137] Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, *et al.* Explainable Aritificial Intelligence (XAI): concepts, taxonomies opportunities and challenges toward responsible AI. *Inf. Fusion* 2020, 58:82–115.