

Artificial Intelligence in Legal Systems: Examining Gender Bias and the Role of UK Legal Frameworks in Addressing It

Muzeng Huang^{1,a,*}

¹*Benenden School, Kent, TN17 4AA, United Kingdom*

a. alinahuang111@gmail.com

**corresponding author*

Abstract: This study examines the gender discrimination of Artificial Intelligence (AI) used in the legal system, focusing on risk assessment, facial recognition, and decision-making and decision-support tools. The study delves into the use of AI in the legal system, examining how its reliance on historical data, under/over-representation, and homogeneity of development teams perpetuate existing gender biases. The study then analyses the implications of the United Kingdom General Data Protection Regulation (UK GDPR) and the proposed Data Protection and Digital Information (DPPI) Bill in addressing gender biases in AI. Nevertheless, the study finds the need for a more robust and proactive legal framework that addresses the root causes of these biases in the design and implementation of AI systems. The paper concludes by proposing a framework to effectively address gender bias in AI systems used in the legal system. The framework outlines explicit obligations across policymakers, companies, and end users to ensure the development and deployment of bias-free AI systems. Its role is to provide comprehensive guidelines and oversight mechanisms that promote proactive measures to prevent gender bias. The framework aims to create a more equitable legal environment for everyone.

Keywords: Artificial Intelligence, Gender Discrimination, UK GDPR, Automated Decision-Making, Policy Recommendations.

1. Introduction

Stereotypes often lead to discrimination in the judiciary, which continues to disadvantage women. Whether as victims, witnesses, or offenders, women's experiences differ significantly from men's [1]. An analysis of 67 million case law documents reveals significant gender bias within the judicial system [2]. With the increasing utilisation of AI in legal systems, will it perpetuate or eliminate gender discrimination? The world has witnessed both its opportunities and risks. AI has been valuable in improving productivity and access to justice, such as through ROSS Intelligence and the DoNotPay System. Legal professionals also believe that using automation in the early stages of court processes is fairer than human judgment, given that gender discrimination is a reality in every judiciary [3]. However, if left unaddressed, AI systems will perpetuate or deepen gender biases, acting as a proxy for human decisions [4].

AI bias stems from two main sources: the use of biased or incomplete datasets for training algorithms and the inherent design biases present within the algorithms themselves [5]. The United Nations Educational, Scientific and Cultural Organization (UNESCO) has identified that large

language models (LLMs), including Llama 2 and GPT-2, exhibit bias against women and girls, a concern intensified by their free and public accessibility [6]. Joy Buolamwini and Timnit Gebru categorised potential harms caused by algorithmic decision-making into three areas: “loss of opportunity,” “economic loss,” and “social stigmatization” [7]. In order to analyse gender discrimination in these domains, this article looks at the effects of automated decision-support and decision-making systems, risk assessment tools, and facial recognition technology.

Next, using the case of the United Kingdom (UK), the paper examines current frameworks and regulations designed to address gender biases in AI systems. Specifically, the UK GDPR aims to protect individuals from potentially harmful legal decisions that are made solely by AI algorithms. Its shortcomings and suggested changes, however, highlight the need for clearer frameworks and preventive mechanisms to address gender discrimination in AI algorithms utilized by the judicial system.

On the whole, this paper utilizes existing literature and case analysis to examine the intersection of AI technology and gender discrimination within the legal system. Then, it critically assesses the UK GDPR and identifies gaps in its effectiveness. By doing so, the study aims to provide valuable insights that stakeholders and policymakers may utilize to create and maintain more equitable AI algorithms for the legal system.

2. Risk Assessment Tools

Organizations use actuarial risk assessment tools to assist judges, prosecutors, and other legal professionals to predict the probability of certain outcomes in court. The risk assessment systems work by analyzing historical datasets and identifying patterns to generate an outcome. However, Katyal underscores the issues of underrepresentation and exclusion, in which certain groups are inadequately represented in certain datasets, leading to inaccurate results [8]. Moreover, risk assessment tools have been male-centric, largely because the majority of violent extremist offenders and terrorists in prison are men [9]. The overreliance on male-centric data leads to inaccurate assessment outcomes for women and other gender identities [9]. Specifically, people often overlook non-binary genders, thereby leading to misclassification.

Moreover, socioeconomic factors, when combined with binary gender variables, also produce biased results. Gwen van Eijim criticises the use of socioeconomic factors in risk assessment tools because they perpetuate social inequality and sentencing disparities. For example, women of socioeconomically marginalised status may be subjected to longer custodial sentences and less favourable treatment in the justice system due to their assessed risk levels [10]. Likewise, Starr articulates the importance of risk assessment tools to focus on individual behaviour and criminal history rather than demographic features. When disproportionately focusing on demographic characteristics, risk assessment tools both fail to yield accurate outcomes for individuals and deepen inequalities and biases within the justice system [11].

Indeed, the U.S. Supreme Court has consistently rejected the use of statistical generalizations about group tendencies, emphasizing individualism as essential to equal protection [12]. For example, in *Craig v. Boren*, the Court ruled against laws that treated individuals differently due to their gender, despite statistical evidence supporting these laws [13].

Given the intention to reject the use of protected variables to yield disparately different predictions, a possible alternative is to explicitly omit gender as a variable to achieve gender-neutral risk assessment tools. However, when gender is not considered a variable nor are gender-specific interpretations taken into account, the predictions become inaccurate. For example, women with a risk score of 6 were found to re-offend at the same rate as men with a risk score of 4 [14]. By enabling risk assessment tools to operate without gender as a variable, they would fail the “calibration within groups” standard and produce unfair predictions [15]. Likewise, a study by Skeem et al. omitted

gender as a factor in the Post Conviction Risk Assessment (PCRA), and they found that the PCRA ended up overestimating the risk of recidivism for women. Therefore, women would be unfairly penalised by risk assessment algorithms if gender was not included as a variable. Afterwards, the authors argue that gender-specific interpretations are necessary in order to produce predictions that are accurate for all [16]. Likewise, Kim's finding on the topic of race-aware algorithms showed that the non-inclusion of race in the risk assessment tools does not ensure fairness, and that it even produces more subtle or exacerbated biases. Kim contends that risk assessment tools should be aware of protected characteristics but should not employ them as decisive factors for prediction outcomes [17].

However, gender-specific risk assessment tools are currently not prevalent: the Radicalization Awareness Network (RAN) claims that the majority of the risk assessment tools now available for violent extremist offenders (VEOs) are majorly designed to evaluate male experiences and behaviours, insufficiently including gender-specific interpretations and indicators. The RAN then emphasizes the importance of incorporating gender-specific factors into existing risk assessment tools in order to understand how identity factors interact and impact experiences. To achieve this goal, the RAN recommends using structured professional judgment (SPJ) for nuanced assessment and including intersectionality [18]. Research shows that a significant percentage of women who are incarcerated report having been victims of physical or sexual abuse before their incarceration [19]. Therefore, rather than letting male-centric risk assessment tools create a cycle of victimization and criminality, gender-specific experiences should be taken into account [19].

Furthermore, the majority of risk assessment tools only take binary gender identities into account, which results in insufficient legal protections for those who do not fit into binary gender classifications [20]. Additionally, this exclusion perpetuates gender stereotypes, which link particular characteristics, behaviours, or appearances to either the male or female category [20].

Thus, neither incorporating gender in a binary form as a risk factor nor omitting it completely is sufficient. On the one hand, risk assessment tools should not rely on a single demographic factor as validation to treat one group disparately differently from another. Furthermore, if current male-centric risk assessment tools are not updated, they will be biased against men due to the extensive data on male offenders, and they will also be inaccurate for women. Overall, risk assessment tools should adopt a comprehensive, intersectional framework to ensure accurate assessments for all gender identities.

3. Facial recognition tools

By analyzing facial features, facial recognition tools enable biometric identification and categorization [21]. Biometric identification requires a database of known faces to match against, allowing police and security agencies to identify suspects and assist in criminal investigations [21]. According to the Government Accountability Office (GAO), seven law enforcement agencies within the Departments of Justice (DOJ) and Homeland Security (DHS) have reported to be utilizing facial recognition technology to aid in criminal investigations [22]. Facial recognition technology functions through a systematic process that includes face detection, feature extraction, and pattern recognition, and it handles variations [23]. Specifically, the program trains an FRT system on a dataset of various faces as it develops. The algorithm learns to distinguish faces from other objects and identifies individual facial features to match new images [24]. This means that the accuracy of these algorithms is heavily dependent on the quality and representativeness of the training data [24].

Joy Buolamwini and Timnit Gebru conducted a study in 2018 that highlighted significant biases in facial recognition technology, revealing lower accuracy rates for women and individuals with darker skin tones. While International Business Machines (IBM), one of the creators of the facial

recognition system, later improved its system and retested it, the error rates still disproportionately affected darker-skinned females [25].

Misidentification caused by insufficiently diverse training datasets can result in false positives, where individuals are wrongly identified as suspects. These false positives can lead to discriminatory treatment and negative experiences [21]. For instance, Porcha Woodruff, a pregnant woman falsely accused of carjacking, was the first known female victim of this phenomenon. She was held for 11 hours and experienced severe physical distress, including a panic attack, and was hospitalized for dehydration after the charges [26]. Unfortunately, this is not a unique case [27].

Furthermore, Schwemmer et al. propose that image recognition systems are an example of the "amplification process," in which they systematically perpetuate existing status inequalities and gender stereotypes, as they categorize men and women with labels that reflect differing statuses [28] [29].

Therefore, despite facial recognition technology's potential to enhance security and aid law enforcement, the biases and inaccuracies inherent in the technology disproportionately affect women by misidentifying them and reinforcing existing status inequalities.

4. Automated decision-support and decision-making

According to the United Kingdom's Information Commissioner's Office (ICO), "Automated decision-making is the process of deciding through automated means without any human involvement" [30]. According to Richardson, automated decision systems refer to any systems, software, or processes that utilize computational methods to assist or substitute for governmental decisions, judgments, and policy execution, affecting opportunities, access, liberties, and/or safety. These systems may include functions such as predicting, classifying, optimizing, identifying, and/or recommending [31].

Nadeem et al. identify three primary sources of bias in AI-based decision-making systems: design and implementation, institutional, and societal [32]. Specifically, a major source of gender bias in AI systems is biased training datasets, which either under-represent or over-represent certain groups [33]. Another mechanism of bias is the lack of gender diversity within AI development teams. This homogeneity fails to account for the experiences and needs of women, reinforcing existing biases in the design and implementation of AI algorithms [34]. Furthermore, the training of AI systems on historical data perpetuates biases due to societal stereotypes and gender roles, which associate certain professions with specific genders [35]. Mimi Onuoha introduced the concept of "algorithmic violence" to describe how automated decision-making systems and algorithms can cause harm by impeding people's access to fundamental needs [36]. Therefore, if unregulated, the use of automated decision support and decision-making will perpetuate gender biases in the legal system.

For example, Amazon developed an unintentionally discriminatory AI recruiting tool in 2014. Since the algorithm was trained on historical hiring data that favoured male candidates, it served as a proxy for humans and discriminated against female applicants for technical roles, perpetuating the existing gender imbalances in the tech industry [37]. This perpetuation shows how algorithmic systems learn and reinforce discriminatory patterns from humans; if used on a wide scale, they would limit women's job opportunities and financial independence. Similarly, Facebook job advertisements favoured male candidates for STEM jobs, credit loan algorithms demonstrated gender bias, and many more [38]. Brooks also highlights demographic biases in custody decision-making algorithms due to historical data, imposing standardized judgments on unique family disputes [39]. This generalization creates a feedback loop: as legal practitioners modify their strategies based on these trends, biases are further entrenched [5].

The increasing integration of AI systems into decision-making processes in the legal system increases the risk of amplifying existing biases, potentially creating a cycle of discrimination [40]. At

the individual level, systematic gender bias leads to unfair treatment, which results in significant economic and social disadvantages for women. These impacts then extend to families to contribute to broader economic inequality, which hinders community development. The significant implications for this risk highlight the importance of addressing gender discrimination in automated decision-making systems.

5. Policy Analysis and Recommendations

5.1. The UK General Data Protection Regulation (UK GDPR)

Currently, the United Kingdom has several legal frameworks and legislation aimed at addressing gender discrimination in AI algorithms. The Equality Act 2010 prohibits discrimination based on protected characteristics, which applies to both human and automated decision-making systems, covering both direct and indirect discrimination [41]. The Human Rights Act 1998 also prohibits discrimination on any grounds [42]. However, both acts are foundational, meaning that they do not explicitly address AI gender discrimination nor lay out clear preventive measures.

The UK General Data Protection Regulation (UK GDPR) is the data privacy and protection law adapted from the European Union General Data Protection Regulation (EU GDPR) after Brexit to suit the UK legal framework. Specifically, Article 22 of the UK GDPR offers safeguards aimed at protecting people from potentially harmful AI decisions that have a legal effect or similarly significant effect on them. Article 22 gives people "the right not to be subject to solely automated decisions, including profiling, which have a legal or similarly significant effect on them." Moreover, Article 22 of the UK GDPR grants individuals the right to contest decisions that have legal or similarly significant effects on them [43]. The article ensures that individuals understand and can interact with decisions that have significant impacts on them. Individuals may also request a human review of AI-driven decisions, which helps to mitigate systematic biases [44]. On the whole, article 22 emphasises communication, accountability, and transparency [43].

Nevertheless, the UK GDPR framework only acts as a reactive mechanism as it addresses biases after they arise rather than preventing them from occurring in the first place. Moreover, the UK GDPR lacks safeguarding explicitly aimed at protecting against gender discrimination in AI algorithms. This means that AI systems in the UK may continue to operate with gender biases without being thoroughly regulated, as long as these systems do not yield a legal or similarly significant impact on individuals. While these biases might not immediately impact legal or similarly significant decisions, they could still perpetuate gender inequality in a data-driven society. Using job ads as an example, Katyal shows how subtle these biases can be and how powerful these subconscious "nudges" can be, even if they do not immediately change behaviour [8]. This subtlety underscores the need for regulatory requirements that proactively address and mitigate the risk of gender bias in AI algorithms.

5.2. The Data Protection and Digital Information Bill

The Data Protection and Digital Information Bill (DPDI) is a legislative proposal that seeks to update the UK's data protection framework. The general aim of the DPDI is to adjust the UK GDPR into a more business-friendly and deregulated framework, which may weaken some of the protections currently offered by the UK GDPR. Below are three main ways the DPDI Bill could influence the protection that individuals receive from UK GDPR.

The first way in which the DPDI Bill could influence the protection that individuals receive from the UK GDPR pertains to the modification of data subject rights. The bill limits the rights of individuals concerning solely automated decision-making, particularly when sensitive data is involved [45]. In situations where the UK GDPR currently prohibits it, the bill also permits solely automated decision-making [46]. Moreover, the DPDI Bill proposes to eliminate the requirement for

a balancing test that weighs the interests of the data controller against the rights of the individual. This change could lead to increased data processing without fully considering the impacts on the data subjects, ultimately weakening their protection [47].

The second way in which the DPDI Bill may affect individual protections relates to the modification of transparency requirements in data processing. The DPDI Bill modifies the requirements for transparency in data processing, particularly concerning Research, Archiving, and Statistical (RAS) purposes [45]. The bill establishes exemptions from proactive transparency requirements as it states "the new derogation when further processing data for research, archiving, and/or statistical purposes" can be applied in situations when "compliance would either be impossible or would involve a disproportionate effort." This implies that if providing the standard transparency information would be too burdensome for an organization, they might not be required to [45].

Additionally, the bill substitutes the standard of "vexatious" requests with the idea of "manifestly unfounded" requests. Organizations now have greater discretion to reject requests that they deem excessively burdensome due to this replacement [45]. Furthermore, the bill narrows the scope of Data Protection Impact Assessments (DPIAs) to require risk documentation just for high-risk processing [45].

The final way that the DPDI Bill might affect individual protection is by changing the integrity obligations that organizations must fulfill. For instance, the DPDI Bill removes the requirement that businesses designate a "senior responsible individual" for high-risk processing activities with the requirement that they designate a statutory Data Protection Officer (DPO) [45].

Additionally, the bill might affect the ICO's independence. This could potentially lead to less effective oversight and enforcement of data protection rights [47].

In general, the DPDI Bill aims to simplify compliance requirements to create flexibility for businesses, particularly medium-sized enterprises (SMEs). Although the government ultimately decided not to proceed with the bill, it signals a future direction in which the UK government attempts to balance innovation with the rights of data subjects.

However, as this trend continues, the crucial step of addressing the root causes of AI gender discrimination remains absent. The future direction of UK legislation should not only create a robust framework to protect against AI bias. Legislation should also develop preventive measures to stop AI biases from arising in the first place. To combat algorithmic bias, this paper recommends the establishment of a robust legal framework that clearly outlines the obligations of policymakers, companies, and users in the design and implementation of bias-free AI systems[48].

5.3. Recommendations

Due to the opacity of AI systems and trade secrets, AI can diminish one's sense of responsibility, deferring everything to the technology. For example, the Estonian initiative that uses AI judges is unclear as to who is responsible for correcting certain errors, whether it is the AI system's developers or the judicial system itself [49].

While the existing legal framework in the United Kingdom lays out a foundation to protect against AI discrimination, it lacks both preventive and reactive mechanisms included in a thorough and explicit framework. To effectively address gender bias in AI systems used in the legal system, the assignment of explicit obligations across different stakeholders is necessary. This proposal is consistent with the European Commission, which aims to introduce a legal framework for AI aimed at defining the responsibilities of users and providers [50]. This framework should ensure that AI systems are developed, deployed, utilized, and feedbacked to prevent bias, promote transparency, and uphold accountability. Below are the proposed obligations for three different stakeholders: policymakers, companies, and users:

5.3.1. Policymakers

Firstly, policymakers need to create specific, bias-aware frameworks for corporations developing AI algorithms used in the judicial system. For example, the UNODC Global Judicial Integrity Network has created award-winning initiatives aimed at eliminating gender bias in AI systems used in legal settings [51]. In order to identify any gender biases in court decisions, policymakers must also fund the creation of an algorithm that makes use of natural language processing (NLP) techniques [48]. These solutions are both feasible, as researchers and organizations have previously developed them. The tools just need to be enhanced and deployed on a wide scale.

Second, legislators must guarantee that judges and other legal professionals are aware of the potential for AI bias and possess a necessary degree of AI literacy. This includes setting up educational programs to train legal professionals in comprehending the interpretations of AI-driven judgments and to be aware of their shortcomings due to the reliance on historical data, forming AI-independent decisions when necessary.

Thirdly, policymakers should implement regulations that explicitly address the domain of gender bias in AI algorithms, as it is currently lacking in all UK frameworks. A UNESCO publication, for example, offers a range of methods for incorporating gender equality into AI principles [52].

Lastly, policymakers should set up independent bias oversight committees responsible for reviewing and addressing biases that stem from AI algorithms used in the judicial system. When end users report bias to these committees, they should review and rectify the cases, and then follow up with the companies that developed the algorithms and impose reasonable penalties as necessary.

5.3.2. Companies

Firstly, companies must comply with government frameworks and regulations to ensure all AI algorithms follow uniform standards and are bias-free and that they are incentivized to make sure they are bias-free.

Besides, companies should be required to conduct bias impact assessments on their AI systems. These assessments identify and address potential biases in the AI systems before they are officially deployed in the judicial system, which prevents any biases from outpouring in the first place. The development process should integrate tools like those developed by Pinton, Sexton, Tozzi, Sevim, and Baker Gillis, which focus on detecting gender biases in legal contexts [48].

Thirdly, companies should adopt various other bias mitigation strategies, such as boxing methods proposed by O'Connor and Liu, which aim to identify and mitigate biases before the full deployment of AI systems [25]. Another effective bias prevention strategy is blind testing, which evaluates AI algorithms across different protected groups to locate biases. If used prior to system deployment, these methodologies would ensure that AI systems are developed with minimal bias.

Fourthly, companies need to set up intermediary explanation pathways that provide clarity regarding AI decisions. This openness makes it possible for litigants and judges to comprehend the reasoning behind AI-driven decisions, allowing them to interact with and respond to them meaningfully.

Fifthly, if biases are found and shown to be caused by an AI system, companies should be held accountable for damages and penalties. This obligation incentivizes businesses to give equity top priority when developing AI systems.

5.3.3. End users

End users include litigants and legal professionals who have a responsibility to use AI algorithms to create an impartial judicial environment. Firstly, users need to be made aware of the potential biases that AI algorithms can create and therefore build independence from solely automated decisions. This

education should include training on how to suspect potential biases and the importance of reporting them. Users need to be incentivized to report and notify bias oversight committees of any biases they identify.

Furthermore, people must be able to fully demand that AI systems be transparent. This includes the freedom to challenge results that seem biased and the right to comprehend the reasoning behind AI decisions. Under the UK GDPR, individuals already have the right to contest solely automated decisions that have a legal or similarly significant effect on them. This right should be expanded to allow users to challenge AI-driven judicial decisions when they suspect any gender bias.

6. Conclusions

On the one hand, AI has the potential to transform the judicial system and society by improving access to justice, efficiency, and consistency. On the other hand, current AI systems have also perpetuated many societal biases. Nevertheless, it is important to acknowledge that AI systems inherently serve as a proxy for human decisions, trying to predict human intent. Gender discrimination would still perpetuate in a world without AI: it is the implementation of AI algorithms in the legal system that has also brought biases to the forefront of human awareness. If AI algorithms deployed in the legal system can be designed and implemented with minimal bias, this awareness could lead to a revolutionary transformation in our society, where justice is free from discrimination. Current UK frameworks, such as the UK GDPR, act as firm reactive measures when individuals are at potential threat under biased automated decisions. However, designing and implementing AI systems with gender equity in mind requires proactive approaches to prevent bias from the outset. Current UK legal frameworks fall short of explicitly addressing gender discrimination in AI applications as well as assigning clear obligations across different stakeholders, which is what this research proposes.

This research is not without its limitations. The focus on frameworks in the UK may not comprehensively reflect regulations and legislations across the globe. Additionally, due to the limitations of up-to-date resources, the research lacks sufficient empirical data and case studies in the judiciary that could illustrate how AI's gender bias impacts the lives of specific individuals, communities, and societies. The proposed framework for addressing gender bias in AI may also not provide detailed implementation strategies or consider the practical challenges of enforcing such measures within specific legal systems. Future research should incorporate specific case studies assessing the impact of gender bias in AI through qualitative or quantitative research. Future studies can also focus on the effectiveness of existing legal frameworks in addressing gender bias in AI and carry out experiments to test the effectiveness and feasibility of proposed frameworks. By addressing these gaps, we can work toward a future where AI serves as a tool for justice free from bias.

References

- [1] *Gender equality*. (2013). https://www.judiciary.uk/wp-content/uploads/JCO/Documents/judicial-college/ETBB_Gender_finalised.pdf
- [2] Baker Gillis, N. (2021, August 1). *Sexism in the Judiciary: The Importance of Bias Definition in NLP and In Our Courts*. *ACLWeb; Association for Computational Linguistics*. <https://doi.org/10.18653/v1/2021.gebnlp-1.6>
- [3] Barysé, D., & Sarel, R. (2023). *Algorithms in the court: does it matter which part of the judicial decision-making is automated?* *Artificial Intelligence and Law*, 32, 117–146. <https://doi.org/10.1007/s10506-022-09343-6>
- [4] Belenguer, L. (2022). *AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry*. *AI and Ethics*, 2(2). <https://doi.org/10.1007/s43681-022-00138-8>
- [5] Zafar, A. (2024). *Balancing the scale: navigating ethical and practical challenges of artificial intelligence (AI) integration in legal practices*. *Discover Artificial Intelligence*, 4(1). <https://doi.org/10.1007/s44163-024-00121-8>
- [6] UNESCO. (2024). *Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes*. *Unesco.org*. <https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes>

- [7] Buolamwini, J., & Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. *Proceedings of Machine Learning Research*, 81(81), 77–91. <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>
- [8] Katyal, S. K. (2020). *Private Accountability in an Age of Artificial Intelligence*. In W. Barfield (Ed.), *The Cambridge Handbook of the Law of Algorithms* (pp. 47–106). chapter, Cambridge: Cambridge University Press.
- [9] Radicalisation Awareness Network. (2023). *The missing gender-dimension in risk assessment Key outcomes*. https://home-affairs.ec.europa.eu/system/files/2024-01/ran_missing_gender-dimension_in_risk_assessment_14112023_en.pdf
- [10] van Eijk, G. (2016). *Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality*. *Punishment & Society*, 19(4), 463–481. <https://doi.org/10.1177/1462474516666282>
- [11] Starr, S. B. (2015). *The New Profiling*. *Federal Sentencing Reporter*, 27(4), 229–236. <https://doi.org/10.1525/fsr.2015.27.4.229>
- [12] Primus, R. A. (2003). *Equal Protection and Disparate Impact: Round Three*. *Harvard Law Review*, 117(2), 493. <https://doi.org/10.2307/3651947>
- [13] US Supreme Court. (1976). *Craig v. Boren*, 429 U.S. 190. *Justia Law*. <https://supreme.justia.com/cases/federal/us/429/190/>
- [14] Drösser, C. (2017, December 22). *In Order Not to Discriminate, We Might Have to Discriminate*. *Simons Institute for the Theory of Computing*. <https://www.droesser.net/en/2017/12/>
- [15] Eckhouse, L., Lum, K., Conti-Cook, C., & Ciccolini, J. (2018). *Layers of Bias: A Unified Approach for Understanding Problems With Risk Assessment*. *Criminal Justice and Behavior*, 46(2), 185–209. <https://doi.org/10.1177/0093854818811379>
- [16] Skeem, J. L., Monahan, J., & Lowenkamp, C. T. (2016). *Gender, Risk Assessment, and Sanctioning: The Cost of Treating Women Like Men*. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2718460>
- [17] Kim, P. (2022, October). *Race-Aware Algorithms: Fairness, Nondiscrimination and Affirmative Action*. *California Law Review*. <https://www.californialawreview.org/print/race-aware-algorithms-fairness-nondiscrimination-and-affirmative-action>
- [18] Directorate-General for Migration and Home Affairs. (2024, May 27). *Improving risk assessment: Accounting for gender*, May 2024. *Migration and Home Affairs*. https://home-affairs.ec.europa.eu/whats-new/publications/improving-risk-assessment-accounting-gender-may-2024_en
- [19] *Women and Girls in the Justice System | Overview*. (2020, August 13). *Office of Justice Programs*. <https://www.ojp.gov/feature/women-and-girls-justice-system/overview#overview>
- [20] Katyal, S., & Jung, J. (2021b). *The Gender Panopticon: Artificial Intelligence, Gender, and Design Justice*. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3760098>
- [21] Waelen, R. A. (2022). *The struggle for recognition in the age of facial recognition technology*. *AI and Ethics*, 3(1). <https://doi.org/10.1007/s43681-022-00146-8>
- [22] Office, U. S. G. A. (2024, March 8). *Facial Recognition Technology: Federal Law Enforcement Agency Efforts Related to Civil Rights and Training | U.S. GAO*. *Www.gao.gov*. <https://www.gao.gov/products/gao-24-107372>
- [23] Lin, S.-H. (2000). *An Introduction to Face Recognition Technology*. *Informing Science: The International Journal of an Emerging Transdiscipline*, 3, 001–007. <https://doi.org/10.28945/569>
- [24] Schuetz, P. (2021). *Fly in the Face of Bias: Algorithmic Bias in Law Enforcement’s Facial Recognition Technology and the Need for an Adaptive Legal Framework*. *Minnesota Journal of Law & Inequality*, 39(1), 221–254. <https://doi.org/10.24926/25730037.626>
- [25] O’Connor, S., & Liu, H. (2023). *Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities*. *AI & SOCIETY*, 39(4), 2045–2057. <https://doi.org/10.1007/s00146-023-01675-4>
- [26] Hill, K. (2023, August 6). *Eight Months Pregnant and Arrested After False Facial Recognition Match*. *The New York Times*. <https://www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest.html>
- [27] Clayton, J. (2024, May 25). *“I was misidentified as shoplifter by facial recognition tech.”* *BBC News*; *BBC News*. <https://www.bbc.co.uk/news/technology-69055945>
- [28] Charles, M. (2012). *Cecilia L. Ridgeway: Framed by Gender: How Gender Inequality Persists in the Modern World*. *European Sociological Review*, 29(2), 408–410. <https://doi.org/10.1093/esr/jcs074>
- [29] Schwemmer, C., Knight, C., Bello-Pardo, E. D., Oklobdzija, S., Schoonvelde, M., & Lockhart, J. W. (2020). *Diagnosing Gender Bias in Image Recognition Systems*. *Socius: Sociological Research for a Dynamic World*, 6(6), 237802312096717. <https://doi.org/10.1177/2378023120967171>
- [30] *What is automated individual decision-making and profiling?* (2023, May 19). *Ico.org.uk*. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/>

- [31] Richardson, R. (2021). *Defining and Demystifying Automated Decision Systems*. *Social Science Research Network*, 81(3).
- [32] Nadeem, A., Marjanovic, O., & Abedin, B. (2022). *Gender bias in AI-based decision-making systems: a systematic literature review*. *Australasian Journal of Information Systems*, 26(26). <https://doi.org/10.3127/ajis.v26i0.3835>
- [33] Veale, M., & Binns, R. (2017). *Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data*. *Big Data & Society*, 4(2), 205395171774353. <https://doi.org/10.1177/2053951717743530>
- [34] Johnson, K. N. (2019, November 14). *Automating the Risk of Bias*. *Ssrn.com*. <https://ssrn.com/abstract=3486723>
- [35] Ntoutsis, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejd, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., & Broelemann, K. (2020). *Bias in Data-driven Artificial Intelligence systems—An Introductory Survey*. *WIREs Data Mining and Knowledge Discovery*, 10(3). <https://doi.org/10.1002/widm.1356>
- [36] Mimi Onuoha . (2021, November 9). *Notes on Algorithmic Violence*. *GitHub*. <https://github.com/MimiOnuoha/On-Algorithmic-Violence>
- [37] Dastin, J. (2018, October 11). *Insight - Amazon scraps secret AI recruiting tool that showed bias against women*. *Reuters*. <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>
- [38] Lambrecht, A., & Tucker, C. (2019). *Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads*. *Management Science*, 65(7), 2966–2981.
- [39] Brooks, W. (2022). *Artificial Bias: The Ethical Concerns of AI-Driven Dispute Resolution in Family Matters*. *Journal of Dispute Resolution*, 2022(2). <https://scholarship.law.missouri.edu/jdr/vol2022/iss2/9>
- [40] Altman, M., Wood, A., & Vayena, E. (2018). *A Harm-Reduction Framework for Algorithmic Fairness*. *IEEE Security & Privacy*, 16(3), 34–45. <https://doi.org/10.1109/msp.2018.2701149>
- [41] GOV.UK. (2010). *Equality Act 2010*. *Legislation.gov.uk; Gov.uk*. <https://www.legislation.gov.uk/ukpga/2010/15/contents>
- [42] *Human Rights Act 1998*. (1998). *Legislation.gov.uk*. <https://www.legislation.gov.uk/ukpga/1998/42/contents>
- [43] ICO. (2023, May 19). *What is the impact of Article 22 of the UK GDPR on fairness?* *Ico.org.uk*. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-is-the-impact-of-article-22-of-the-uk-gdpr-on-fairness/>
- [44] ICO. (2023, May 19). *What about fairness, bias and discrimination?* *Ico.org.uk*. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/what-about-fairness-bias-and-discrimination/>
- [45] Erdos, D. (2022). *A Bill for a Change? Analysing the UK Government’s Statutory Proposals on the Content of Data Protection and Electronic Privacy*. *SSRN Electronic Journal*, 13. <https://doi.org/10.2139/ssrn.4212420>
- [46] *How the new Data Bill waters down protections*. (2023, November 28). *Public Law Project*. <https://publiclawproject.org.uk/resources/how-the-new-data-bill-waters-down-protections/>
- [47] McCullagh, K. (2023). *Data Protection and Digital Sovereignty Post-Brexit*. *Bloomsburycollections.com*. <https://www.bloomsburycollections.com/monograph-detail?docid=b-9781509966516&tocid=b-9781509966516-chapter2>
- [48] Raysa Benatti, Severi, F., Avila, S., & Colombini, E. L. (2024). *Gender Bias Detection in Court Decisions: A Brazilian Case Study*. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 67(3), 746–763. <https://doi.org/10.1145/3630106.3658937>
- [49] Bell, F., Bennett Moses, L., Legg, M., Silove, J., & Zalnieriute, M. (2022, June 14). *AI Decision-Making and the Courts: A Guide for Judges, Tribunal Members and Court Administrators*. *Papers.ssrn.com*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4162985
- [50] Di Noia, T., Tintarev, N., Fatourou, P., & Schedl, M. (2022). *Recommender systems under European AI regulations*. *Communications of the ACM*, 65(4), 69–73. <https://doi.org/10.1145/3512728>
- [51] *Award-winning project on preventing gender bias in AI systems used in judiciaries*. (2021). *United Nations : Office on Drugs and Crime*. <https://www.unodc.org/unodc/en/gender/news/award-winning-project-on-preventing-gender-bias-in-ai-systems-used-in-judiciaries.html>
- [52] UNESCO. (2020). *Artificial intelligence and gender equality: Key findings of UNESCO’s global dialogue*. <https://unesdoc.unesco.org/ark:/48223/pf0000374174>