Title: Evaluating the Effectiveness of Anti-Discrimination Laws of EU and UK in Mitigating Biases in AI Automated Decision-Making Systems: Legal Reforms and Future Directions

Author: Stella Isimbi

**Contents**

[SCHOOL OF LAW 1](#_Toc177050640)

[**COURSEWORK COVER SHEET (2023-24)** 1](#_Toc177050641)

[ASSIGNMENT DETAILS 1](#_Toc177050642)

[WORD COUNT 1](#_Toc177050643)

[SUBMISSION DEADLINE 2](#_Toc177050644)

[UNIVERSITY REGULATIONS 2](#_Toc177050645)

[**Acknowledgements** 2](#_Toc177050646)

[**Chapter 1: ABSTRACT** 3](#_Toc177050647)

[**Chapter 2: INTRODUCTION** 3](#_Toc177050648)

[**2.1 Background and Context** 3](#_Toc177050649)

[**2.2 Research Problem** 4](#_Toc177050650)

[**2.3 Research Questions** 4](#_Toc177050651)

[**2.4 Methodology** 4](#_Toc177050652)

[**2.5 Structure of the Thesis** 4](#_Toc177050653)

[**2.6 Significance of the Study** 5](#_Toc177050654)

[**Chapter 3: AI Automated Decision-Making Systems: An Overview** 5](#_Toc177050655)

[**3.1 Definition and Mechanisms of AI Automated Decision-Making Systems** 5](#_Toc177050656)

[**3.2 Common Areas of Application** 6](#_Toc177050657)

[**3.3 Bias in AI Systems** 6](#_Toc177050658)

[**3.4 Implications of AI Bias in Decision-Making** 7](#_Toc177050659)

[**3.5 Addressing Bias in AI: Current Approaches** 11](#_Toc177050660)

[**Chapter 4: Anti-Discrimination Laws in the EU and UK: A Critical Review** 13](#_Toc177050661)

[**4.1 Historical Overview of Anti-Discrimination Laws** 13](#_Toc177050662)

[**4.2 Key Legal Instruments** 13](#_Toc177050663)

[**4.4 Conclusion** 19](#_Toc177050664)

[**Chapter 5: Evaluating the Effectiveness of Current Legal Frameworks** 20](#_Toc177050665)

[**5.1 Introduction** 20](#_Toc177050666)

[**5.2 Strengths of Existing Legal Frameworks** 20](#_Toc177050667)

[**5.3 Weaknesses and Gaps in the Legal Frameworks** 22](#_Toc177050668)

[**5.5 Conclusion** 25](#_Toc177050669)

[**Chapter 6: Legal Reforms and Future Directions** 25](#_Toc177050670)

[**6.1 Introduction** 25](#_Toc177050671)

[**6.2 Enhancing AI Transparency and Accountability** 26](#_Toc177050672)

[**6.3 Strengthening Data Protection and Bias Mitigation** 27](#_Toc177050673)

[**6.4 Developing AI-Specific Anti-Discrimination Legislation** 28](#_Toc177050674)

[**6.5 International Collaboration and Harmonization of Standards** 28](#_Toc177050675)

[**6.6 Conclusion** 29](#_Toc177050676)

[**Chapter 7: Conclusion** 29](#_Toc177050677)

[**7.1 Summary of Findings** 29](#_Toc177050678)

[**7.2 Implications for Policy and Practice** 30](#_Toc177050679)

[**7.4 Final Thoughts** 31](#_Toc177050680)

[**References** 32](#_Toc177050681)

## **Acknowledgements**

First and foremost, I would like to thank God for His unwavering guidance, strength, and blessings throughout this journey. Without His grace, none of this would have been possible. I am deeply grateful for the wisdom and resilience He has provided, enabling me to persevere and complete this work.

And to my amazing Parents, whose unwavering love, sacrifices, support and belief in me have been the bedrock of my success. Your constant support has given me the strength to pursue my goals, I am so blessed for having you. To my siblings, thank you for always being there, offering encouragement, laughter, and the occasional push when needed.

I would also like to express my heartfelt thanks to my supervisor, Dr Eugen Lim, for his constant support, insightful feedback, and invaluable guidance. Your expertise and patience have been instrumental in shaping this research, and I am incredibly fortunate to have had you as a mentor.

To Dr Ciarán O'Kelly and Professor Giancarlo Frosio, thank you for your inspiration and dedication. Your teachings have profoundly impacted my understanding, and your encouragement has been a source of motivation throughout my studies.

This accomplishment reflects the love, support, and guidance I have received from each of you. Thank you.**Top of Form**

## **Bottom of Form**

## **Chapter 1: ABSTRACT**

This thesis argues that current legal frameworks in the EU and UK are insufficient in mitigating AI bias, and proposes targeted legal reforms, including AI-specific anti-discrimination legislation and enhanced transparency requirements.

Through a detailed exploration of the existing legal landscape in the EU and UK, this research critically assesses whether current regulations are truly safeguarding individual rights and promoting fairness. By drawing on comprehensive literature reviews, case studies, and legal precedents, the thesis identifies significant gaps where protections fall short.

## **Chapter 2: INTRODUCTION**

### **2.1 Background and Context**

The rapid advancement and adoption of Artificial Intelligence (AI) technologies are transforming decision-making processes across a variety of sectors, from finance and healthcare to law enforcement and human resources. AI-driven Automated Decision-Making Systems (ADMS) are increasingly tasked with making critical decisions, such as loan approvals, hiring, medical diagnoses, and even determining criminal sentences. [[1]](#footnote-1) These systems promise greater efficiency, consistency, and speed compared to traditional human-led processes. However, with these advancements come significant ethical and legal concerns, particularly around fairness, accountability, and transparency.

One of the most pressing concerns is the potential for AI systems to perpetuate, or even exacerbate, existing societal biases. The data used to train these systems often reflects historical inequalities, such as gender, racial, and socio-economic biases. As AI systems learn from this data, they risk replicating discriminatory patterns in their decision-making processes. Instances of AI-driven recruitment tools that disadvantage women and minorities,[[2]](#footnote-2) and predictive policing systems that disproportionately target communities of colour,[[3]](#footnote-3) underscore the real-world impact of these biases.

The intersection of AI and the law presents new challenges. Traditional anti-discrimination laws, designed to address human bias, struggle to adapt to the opaque and complex nature of AI systems. Moreover, the legal frameworks in the European Union (EU) and the United Kingdom (UK) are not fully equipped to address the intricacies of algorithmic decision-making. While regulations such as the General Data Protection Regulation (GDPR) in the EU and the Equality Act 2010 in the UK offer some protections, significant gaps remain in their application to AI systems.

### **2.2 Research Problem**

The deployment of AI in decision-making processes has raised critical questions about the adequacy of existing anti-discrimination laws in the EU and UK. Can these legal frameworks, designed to combat human bias, adequately address the biases embedded in AI systems? Furthermore, the opacity of many AI systems, often referred to as the "black box" problem, makes it difficult for individuals to challenge biased decisions or understand how these decisions are made.

This thesis investigates whether current anti-discrimination laws in the EU and UK are effective in mitigating biases in AI-driven ADMS. It critically examines whether legal reforms are needed to ensure that AI systems promote fairness, equality, and transparency. The research not only assesses the strengths and weaknesses of existing laws but also proposes actionable reforms to bridge the gaps in legal protection.

### **2.3 Research Questions**

This research is guided by the following key questions:

* How effective are current EU and UK anti-discrimination laws in addressing biases in AI Automated Decision-Making Systems (ADMS)??
* What gaps exist within these legal frameworks that may allow AI-induced discrimination to persist??
* What legal reforms or policy measures are necessary to enhance the fairness and accountability of AI-driven decision-making systems?

### **2.4 Methodology**

To explore these questions, this study employs a qualitative research approach combining legal analysis with a review of academic literature and case studies. It critically examines the existing legal frameworks in the EU and UK through the following methods:

* **Legal Analysis**: A critical examination of anti-discrimination laws, such as the GDPR, the Equality Act 2010, and the proposed EU AI Act, focusing on their application to AI-driven systems.
* **Literature Review**: A review of academic and policy discussions on AI bias and the law, drawing on interdisciplinary sources from fields such as law, ethics, and computer science.
* **Case Studies**: Detailed analysis of significant legal cases where AI systems have been challenged for bias, offering insights into the judicial interpretation of anti-discrimination laws in both the EU and UK.

### **2.5 Structure of the Thesis**

This thesis is organized into seven chapters. Chapter 1 provides an abstract, while Chapter 2 (this chapter) outlines the introduction and research context. Chapter 3 offers a detailed exploration of AI Automated Decision-Making Systems, explaining how biases can be embedded in these systems and the potential societal implications. Chapter 4 reviews the current anti-discrimination laws in the EU and UK, assessing their application to AI technologies. Chapter 5 evaluates the effectiveness of these legal frameworks and identifies gaps and weaknesses. Chapter 6 proposes legal reforms and explores future directions for mitigating AI bias through enhanced legal protections. Finally, Chapter 7 summarizes the key findings, discusses the policy implications, and suggests avenues for future research.

### **2.6 Significance of the Study**

This study addresses an urgent gap in the intersection of law and technology. As AI systems continue to influence high-stakes decisions, the potential for unchecked biases poses significant risks to fairness, equality, and justice. By evaluating the effectiveness of current anti-discrimination laws and proposing legal reforms, this research aims to inform policymakers and contribute to the development of AI governance frameworks that prioritize human rights. The findings of this study have the potential to shape legal and policy debates on AI, ensuring that technological advancements do not come at the cost of societal equity.

As we move forward, Chapter 3 will provide an in-depth overview of AI automated decision-making systems, exploring the mechanisms through which they operate and the potential biases that can emerge in their design and application. This exploration will lay the groundwork for understanding the legal responses required to mitigate bias in AI-driven decision-making processes.

## **Chapter 3: AI Automated Decision-Making Systems: An Overview**

### **3.1 Definition and Mechanisms of AI Automated Decision-Making Systems**

Artificial Intelligence (AI) automated decision-making systems (ADMS) are sophisticated computational models designed to perform tasks that traditionally required human intelligence. These systems utilize algorithms, particularly machine learning (ML) models, to analyse vast datasets, identify patterns, and make decisions or predictions based on that data. The automated decision-making process in AI systems can range from relatively simple rule-based systems to complex neural networks that mimic human cognitive functions.[[4]](#footnote-4)

At the core of AI ADM is the ability to process large amounts of data with speed and accuracy far beyond human capabilities. These systems often employ techniques such as supervised learning, where models are trained on labelled data; unsupervised learning, where models identify patterns without labelled outcomes; and reinforcement learning, where systems learn by receiving feedback on their actions.[[5]](#footnote-5) As a result, AI ADM systems have been integrated into various sectors, playing a critical role in shaping outcomes in areas like finance, healthcare, law enforcement, and human resources.

### **3.2 Common Areas of Application**

The application of AI in ADM processes spans multiple sectors, each benefiting from the efficiencies and insights that AI ADM systems offer. Below are some of the key areas where AI is commonly utilized:

In the financial sector, AI ADM systems are employed for tasks such as credit scoring, fraud detection, and investment management. For instance, AI-driven algorithms assess loan applications by analysing an applicant’s credit history, spending patterns, and other financial behaviours. However, these systems can unintentionally perpetuate historical biases, such as those based on race or socioeconomic status, leading to discriminatory lending practices.[[6]](#footnote-6)

Additionally, AI ADM systems in healthcare are used for diagnosing diseases, personalizing treatment plans, and managing patient care. Machine learning models can analyse medical records and imaging data to predict patient outcomes. Despite their potential, there are concerns about the accuracy and fairness of AI-driven diagnoses, especially when the training data does not adequately represent diverse populations.[[7]](#footnote-7)

In addition, Predictive policing tools use AI to forecast criminal activity based on historical crime data. These systems aim to optimize the allocation of law enforcement resources. However, they have been criticized for reinforcing existing biases in policing, particularly against minority communities, by predicting higher crime rates in already over-policed areas.[[8]](#footnote-8)

And finally,AI-powered recruitment tools such as Applicant Tracking System (ATS) streamline the hiring process by screening resumes, conducting initial interviews, and even predicting candidate success based on previous hires. Yet, these systems can inherit biases present in historical hiring data, leading to the exclusion of qualified candidates from underrepresented groups.[[9]](#footnote-9)

### **3.3 Bias in AI Systems**

Bias in AI systems remain a significant concern, particularly in high-stakes decision-making areas like healthcare, criminal justice, and employment. Bias can arise from multiple sources, including data quality, algorithmic design, and deployment context. In each of these cases, AI systems often replicate historical patterns of discrimination, leading to unfair outcomes for certain groups. The following examples highlight how bias has manifested in real-world AI applications and illustrate the broad implications of biased AI systems.

#### **3.3.1 Employment: The Case of Amazon’s AI Recruiting Tool**

A notable instance of AI bias occurred with Amazon’s AI-powered recruiting tool, discontinued in 2018 after it was found to disadvantage female applicants. Trained on resumes from a male-dominated workforce, the AI *favoured* male candidates and penalized resumes mentioning terms like "women’s" or female-specific programs. This case illustrates data bias: if the training data reflects societal inequalities, the AI system will likely reproduce these biases unless corrective measures, such as using more representative data or fairness-aware algorithms, are implemented

#### **3.3.2 Criminal Justice: Predictive Policing Systems**

In the criminal justice sector, predictive policing algorithms have gained popularity in recent years. These systems, like PredPol (Predictive Policing), use historical crime data to forecast where crimes are likely to occur and guide law enforcement in resource allocation. However, research has shown that these systems disproportionately target minority communities, particularly African American and Hispanic neighbourhoods, due to biases embedded in historical crime data.

For example, in jurisdictions like Chicago and Los Angeles, predictive policing systems were found to direct more police resources toward predominantly Black and Hispanic communities, reinforcing existing patterns of over-policing. The data used to train these models often reflects years of biased law enforcement practices,[[10]](#footnote-10) such as racial profiling and unequal policing efforts. As a result, the algorithms "learn" to focus on these communities, which not only amplifies discriminatory outcomes but also perpetuates a cycle of surveillance and arrest that disproportionately affects people of colour.

This case illustrates contextual bias, where the deployment context of the AI system (i.e., the historical inequalities in law enforcement) exacerbates biased outcomes.[[11]](#footnote-11) It also underscores the "black box" nature of many AI systems, where the decision-making process is opaque, making it difficult for affected individuals to challenge the AI's conclusions.

#### **3.3.3 Healthcare: Racial Bias in Medical Algorithms**

AI in healthcare has the potential to enhance diagnostics and treatment, but studies show that many algorithms exhibit racial biases, leading to unequal treatment for minority patients. For instance, a 2019 study revealed that an AI algorithm used to predict patients’ need for additional healthcare resources, impacting over 200 million people annually in the U.S., showed a bias against Black patients. The algorithm relied on healthcare spending as a proxy for health needs, leading to under-treatment of Black patients due to historically lower healthcare expenditures. This example underscores the need for unbiased and appropriate variables in AI models, especially in critical areas like healthcare, where biased outcomes can have severe consequences.

#### **3.3.4 Facial Recognition: Misidentification of Minorities**

Facial recognition technology, used in various sectors, has been shown to perform poorly for certain demographics. A 2018 MIT Media Lab study found that facial recognition algorithms were significantly less accurate in identifying women and people with darker skin tones, with error rates for darker-skinned women reaching 34% compared to under 1% for lighter-skinned men.

In 2020, Robert Williams, an African American man, was wrongfully arrested in Detroit after a facial recognition system misidentified him as a suspect. This case highlights technical bias in facial recognition, where inadequate training data, often skewed towards lighter skin tones, results in poor accuracy for minority populations. It underscores the serious risks of biased AI in law enforcement, where errors can lead to significant civil rights violations.

In conclusion, Bias in AI can exacerbate existing inequalities and lead to discriminatory outcomes across multiple sectors, from denying someone a job or medical treatment to wrongfully targeting individuals for arrest. It is clear that addressing AI bias requires not only technological solutions, such as improving data quality and algorithmic fairness but also robust legal frameworks that ensure accountability, transparency, and fairness in AI-driven decision-making systems.

### **3.4 Implications of AI Bias in Decision-Making**

As AI technologies become increasingly integrated into high-stakes domains, the consequences of biased decision-making can no longer be viewed as purely technical or operational issues. They have deep social, ethical, and legal implications, which can affect individuals and groups at a systemic level. Below are some of the key implications that arise from bias in AI-driven decision-making systems.

#### **3.4.1 Discrimination**

One of the most immediate and troubling implications of bias in AI decision-making systems is the perpetuation of discriminatory outcomes, which directly violate core principles of fairness and equality. AI systems that are designed to automate decisions in crucial areas are often trained on data that reflects historical inequalities or societal biases. As a result, these systems can replicate or even amplify existing prejudices, leading to outcomes that disproportionately affect certain groups.

For instance, biased recruitment algorithms can entrench gender or racial disparities in the workplace by favouring candidates who resemble past hires, who may have been predominantly male or from a specific racial background. Research has shown that certain AI hiring tools have systematically disadvantaged women and minority candidates by ranking them lower than their white, male counterparts, often because the training data reflected past hiring practices that favoured these groups.[[12]](#footnote-12) Similarly, biased credit scoring systems, which rely on historical data about borrowers' creditworthiness, can result in unequal access to financial services for minority groups. A system that associates higher default rates with certain racial or socio-economic groups may deny loans to individuals from these communities, even if they present a low financial risk.[[13]](#footnote-13) These discriminatory outcomes reinforce systemic inequalities and deny individuals equal opportunities, in clear contravention of anti-discrimination laws and human rights protections.

Additionally, AI decision-making systems depend heavily on public trust for their widespread adoption and use. However, when AI systems are perceived as biased or discriminatory, they undermine that trust, leading to scepticism and resistance from the public.

In healthcare, for instance, AI algorithms are being developed to assist with diagnosis, treatment recommendations, and resource allocation. If these systems are found to be biased, such as underdiagnosing conditions in certain demographic groups, like women or people of colour the trust that patients and healthcare professionals place in AI-driven solutions can erode rapidly.[[14]](#footnote-14) Without trust, patients may be reluctant to accept AI-driven medical advice, which in turn limits the technology's ability to improve healthcare outcomes. Similarly, in finance, biased algorithms that deny loans or increase interest rates for marginalized groups can lead to public outrage, legal challenges, and a hesitance to use AI-driven financial tools.

As a result, the erosion of trust can affect the broader adoption of AI technologies. When the public perceives AI as inherently unfair or biased. therefore, risks might arise because of this scepticism and can extend to other sectors, stalling innovation and slowing the implementation of potentially beneficial AI systems. Building and maintaining trust, therefore, requires a concerted effort to ensure that AI systems are transparent, accountable, and demonstrably fair in their decision-making processes.[[15]](#footnote-15)

#### **Legal and Regulatory Challenges**

The discriminatory impacts of biased AI systems present significant legal and regulatory challenges, particularly in jurisdictions with strong anti-discrimination laws such as the **European Union (EU)** and the **United Kingdom (UK).** As AI systems are increasingly used in important sectors, concerns about fairness and equality are heightened. As a result, Organizations deploying AI systems are not exempt from legal obligations to ensure equality; in fact, the complexity and opacity of AI may subject these organizations to **heightened scrutiny** from regulators and expose them to substantial legal liabilities if their systems are found to produce discriminatory outcomes.

In the EU, the **General Data Protection Regulation (GDPR)** serves as a critical legal framework governing the use of automated decision-making systems, particularly when such systems affect individuals’ rights and freedoms. **Article 22** of the GDPR specifically addresses automated decision-making, including profiling, and provides individuals with the right not to be subject to decisions made solely by automated processes that have significant legal effects, unless certain conditions are met. These conditions include obtaining explicit consent from the individual or the need to fulfil a contract. The provision is especially relevant for AI systems used in high-stakes contexts such as recruitment, lending, and law enforcement, where biased algorithms can have serious repercussions on individuals' lives[[16]](#footnote-16).

The GDPR goes beyond simply requiring transparency; it demands accountability, particularly when personal data is involved in automated decisions. AI systems that make decisions without human intervention and produce biased outcomes such as denying loans to individuals from minority ethnic groups or disproportionately targeting certain demographics in predictive policing can lead to **legal challenges under GDPR**. Therefore, A critical aspect of GDPR enforcement lies in the notion of **data protection impact assessments (DPIAs)**, which organizations are required to conduct before deploying high-risk AI systems. These assessments are designed to identify and mitigate risks, including the risk of discrimination. Failure to adequately assess these risks or prevent discriminatory outcomes can result in significant penalties under GDPR, ranging from financial fines to reputational damage for the organization.

Moreover, the GDPR imposes specific limitations on the processing of **sensitive personal data**, such as racial or ethnic origin, political opinions, and health information, under **Article 9**. AI systems that process such data for decision-making purposes must meet stringent requirements, including demonstrating that the processing is necessary and obtaining explicit consent from the individuals involved. These provisions are crucial in preventing discriminatory practices, especially in areas like recruitment, where AI algorithms can unintentionally replicate historical biases if trained on biased datasets.

However, while GDPR provides strong protections against discriminatory automated decision-making, enforcement challenges remain. One of the key criticisms of the GDPR is that the **right to explanation**, embedded in Article 22, is not always clear. AI systems, particularly those relying on complex machine learning algorithms, can be difficult to interpret even for their developers. This opacity, often referred to as the **black box problem**, complicates efforts to ensure transparency and accountability. Individuals subject to AI decisions may not always receive adequate explanations for adverse outcomes, making it difficult to challenge those decisions under GDPR. Therefore, the regulation’s effectiveness in addressing AI bias depends largely on how well it is enforced and how courts interpret the right to explanation in the context of AI.

In contrast, the **United Kingdom (UK)** approaches the regulation of AI systems primarily through the **Equality Act 2010**, which prohibits direct and indirect discrimination on the basis of protected characteristics, including race, gender, age, disability, and religion. The Equality Act holds organizations liable for discriminatory practices, regardless of intent. This is particularly important in the context of AI, where algorithms may inadvertently produce biased outcomes without any deliberate intent to discriminate. For example, an AI system designed to screen job applicants may prioritize candidates who resemble past successful employees, thereby disadvantaging women or minority candidates if those groups were historically underrepresented in the organization’s workforce[[17]](#footnote-17).

The **indirect discrimination** provisions of the Equality Act are particularly relevant in the AI context. Indirect discrimination occurs when a seemingly neutral policy or practice has a disproportionately negative impact on individuals with certain protected characteristics. AI systems, which rely on historical data and statistical models, can easily reproduce existing inequalities if not properly designed and monitored. For example, the UK welfare system’s use of AI in benefits assessments was challenged in the case of **R (on the application of Edward Santon) v Secretary of State for Work and Pensions**, where the court found that the automated system failed to account for the specific needs of disabled claimants, leading to indirect discrimination under the Equality Act[[18]](#footnote-18).

AI bias has also surfaced in **law enforcement** practices in the UK, where automated systems like facial recognition technology have been used to identify individuals in public spaces. The case of **R (on the application of Bridges) v Chief Constable of South Wales Police** highlighted the discriminatory potential of AI in policing. The court found that the deployment of facial recognition technology violated privacy and anti-discrimination laws, particularly due to its higher rates of misidentification for individuals from minority ethnic groups. The case demonstrates how AI systems in law enforcement can perpetuate racial disparities, raising serious concerns about fairness and equality in criminal justice decision-making[[19]](#footnote-19)

The UK’s **regulatory framework** faces similar challenges as the GDPR in ensuring that AI systems are transparent and accountable. While the Equality Act provides a robust legal foundation for challenging discriminatory outcomes, courts are increasingly grappling with the complexities of AI-driven decision-making. One of the critical issues is the lack of transparency in how AI algorithms reach their conclusions, which makes it difficult for individuals to contest decisions or prove discrimination. This challenge is compounded by the fact that AI systems are often proprietary and protected by intellectual property laws, limiting the ability of affected individuals or regulators to scrutinize the decision-making processes fully.

The growing number of legal challenges against biased AI systems in both the EU and UK underscores the need for **clearer regulatory guidelines** and more effective enforcement mechanisms. Courts and regulators must balance the benefits of AI innovation with the fundamental rights to equality and fairness. In both jurisdictions, **human oversight** is a critical component in mitigating the risks of AI bias. Requiring human intervention in high-risk decision-making processes can serve as a safeguard against the unchecked power of algorithms. However, the success of these measures ultimately depends on the willingness of organizations to prioritize fairness in the design and deployment of AI systems, and the ability of regulatory bodies to hold them accountable.

Beyond legal and regulatory issues, AI bias also has profound societal and ethical implications. AI systems that produce biased outcomes can deepen existing social inequalities and exacerbate systemic injustices. For example, in criminal justice, AI tools used for predictive policing have been criticized for disproportionately targeting minority communities based on biased historical crime data. Such tools can perpetuate cycles of over-policing and surveillance in these communities, exacerbating already entrenched disparities in the criminal justice system.[[20]](#footnote-20)

Ethically, there is a growing concern about the use of AI in decision-making without adequate safeguards to prevent harm. Philosophers and ethicists argue that fairness and justice should be at the core of any AI system that impacts human lives. The deployment of biased AI systems, particularly in critical sectors like healthcare and criminal justice, raises questions about whether we are willing to accept automated decisions that may lead to unjust outcomes for certain groups. As AI systems increasingly assume roles that directly impact human rights, the ethical responsibility to mitigate bias and ensure fairness becomes even more urgent.[[21]](#footnote-21)

Finally, the economic implications of biased AI systems are not to be overlooked. Discriminatory AI systems can have negative financial consequences for both individuals and organizations. For individuals, being unfairly denied a loan, job, or medical treatment due to a biased AI decision can have long-lasting financial repercussions, deepening economic inequalities. For organizations, biased AI systems can lead to costly legal battles, regulatory fines, and reputational damage. Furthermore, failing to address bias in AI systems can stifle innovation, as companies that do not prioritize fairness and inclusivity in their AI development may struggle to gain public trust or comply with increasingly stringent regulatory requirements.[[22]](#footnote-22)

#### **3.4.6 Conclusion**

The implications of bias in AI decision-making systems are far-reaching, affecting not only individual rights and freedoms but also the broader societal trust in AI technologies and the organizations that deploy them. As AI becomes more integral to decision-making processes, ensuring that these systems are fair, transparent, and accountable is essential. Addressing bias is not only a legal requirement but also a moral obligation that impacts the legitimacy of AI systems and their role in shaping the future of society.

### **3.5 Addressing Bias in AI: Current Approaches**

The challenge of mitigating bias in AI ADM systems has become a focal point in both technological development and ethical governance. As AI systems are increasingly integrated into decision-making processes that affect individuals’ lives whether in hiring, healthcare, lending, or law enforcement ensuring these systems operate fairly is not only a legal obligation but also a moral imperative. To address these concerns, several promising strategies have been proposed and are gradually being implemented across sectors. These approaches reflect the growing recognition that both technical innovation and ethical oversight are essential to preventing AI systems from perpetuating or exacerbating existing inequalities.

**Fairness-Aware Algorithms**

One of the most significant advancements in this field has been the development of fairness-aware algorithms. Researchers are designing these algorithms to optimize for fairness alongside traditional performance metrics like accuracy. Rather than simply producing the most statistically accurate predictions, these algorithms are crafted to adjust decision-making processes in ways that reduce or eliminate biases. By explicitly incorporating fairness constraints into their models, developers aim to ensure more equitable outcomes across different demographic groups. This means, for instance, ensuring that an algorithm used in hiring does not disproportionately disadvantage women or minority candidates, or that a lending algorithm is not skewed against lower-income individuals. These fairness-aware algorithms represent a proactive approach to embedding ethical considerations directly into the design of AI systems, addressing bias at the source.[[23]](#footnote-23)

**Algorithmic Audits**

Algorithmic audits have emerged as a crucial tool in the effort to mitigate bias in AI systems. These audits involve a comprehensive examination of an AI system’s inputs, outputs, and underlying decision-making processes. The goal is to detect potential biases and ensure that the system is compliant with ethical standards and legal obligations. Regular audits can reveal hidden biases that may not be apparent in initial system testing, particularly as AI models evolve over time with new data inputs. By conducting ongoing audits, organizations can identify discrepancies between expected and actual system behavior, ensuring that bias does not creep into the system’s operations over time. More than a technical exercise, algorithmic audits also promote a culture of transparency and accountability, reinforcing the notion that those who design and deploy AI systems must be vigilant in safeguarding against discriminatory outcomes.[[24]](#footnote-24)

**Diverse Data Sets**

Another critical approach to mitigating bias in AI is ensuring that the data used to train these systems is diverse and representative of the populations affected by the decisions. AI models learn from data, and if the data reflects historical inequalities or underrepresents certain groups, the models are likely to reproduce these patterns in their outcomes. Ensuring diversity in data collection means gathering data from a broad range of demographic groups, particularly those who have historically been marginalized or underrepresented. By doing so, developers can reduce the risk that AI systems will disproportionately harm or exclude certain groups. Importantly, this process requires not only technical expertise but also a deep understanding of the social contexts in which these systems operate. Ethical considerations must guide the collection and use of data to ensure that the pursuit of diversity does not inadvertently violate privacy or other rights.[[25]](#footnote-25)

**Transparency and Accountability**

Transparency and accountability are foundational to addressing bias in AI systems. Without transparency, it becomes impossible to understand how an AI system is making its decisions, let alone identify whether those decisions are biased. Enhancing transparency involves making the decision-making processes of AI systems more understandable to those who interact with or are affected by them. This may include providing clear explanations of how specific decisions were reached or offering insight into the criteria used by the algorithm. Transparency empowers users, regulators, and affected individuals to scrutinize AI systems more effectively, creating a feedback loop that encourages fairness.

Equally important is the need for accountability in AI deployment. Establishing clear lines of responsibility ensures that organizations and developers are held accountable for the outcomes of their AI systems. This means that if an AI system produces biased or discriminatory results, there are mechanisms in place to identify the responsible parties and hold them to account. Accountability mechanisms can include legal remedies, such as the right to contest decisions, as well as internal organizational policies that ensure ethical oversight and compliance with both legal and ethical standards.[[26]](#footnote-26)

Together, these approaches fairness-aware algorithms, algorithmic audits, diverse data sets, and transparency and accountability represent a multifaceted strategy for addressing bias in AI. However, none of these solutions can work in isolation. A comprehensive and holistic approach, combining technological innovation with ethical governance, is essential to ensuring that AI systems serve the interests of all individuals, particularly those who have been historically marginalized. As AI continues to evolve and shape critical decisions in society, the commitment to fairness, transparency, and accountability must remain at the forefront of its development.

## **Chapter 4: Anti-Discrimination Laws in the EU and UK: A Critical Review**

### **4.1 Historical Overview of Anti-Discrimination Laws**

The evolution of anti-discrimination laws in the European Union (EU) and the United Kingdom (UK) has been shaped by a commitment to equality, human rights, and the rule of law. These legal frameworks have sought to address systemic inequalities and protect individuals from discrimination on various grounds, including race, gender, age, disability, religion, and sexual orientation.

In the EU, anti-discrimination laws have their roots in the foundational treaties, particularly the Treaty of Rome (1957), which established the European Economic Community and laid the groundwork for non-discrimination principles in employment and social policies. The subsequent development of directives, such as the Equal Treatment Directive (1976) and the Racial Equality Directive (2000), further strengthened the EU’s commitment to combating discrimination.[[27]](#footnote-27) The Charter of Fundamental Rights of the European Union (2000), which became legally binding with the Treaty of Lisbon (2009), enshrines the right to non-discrimination in Article 21, providing a robust legal basis for challenging discriminatory practices across member states.[[28]](#footnote-28)

On the hand, In the UK, anti-discrimination laws have evolved through a combination of domestic legislation and EU directives, which were incorporated into UK law. The Race Relations Act 1965 was the first significant anti-discrimination law in the UK, prohibiting discrimination on the grounds of race in public places. This was followed by the Sex Discrimination Act 1975, the Disability Discrimination Act 1995, and other legislation aimed at protecting various groups from discrimination. The Equality Act 2010 consolidated and harmonized these laws, providing a comprehensive legal framework to combat discrimination in the UK.[[29]](#footnote-29)

### **4.2 Key Legal Instruments**

The current legal frameworks in the EU and UK are underpinned by several key legal instruments designed to protect individuals from discrimination. These instruments have been instrumental in shaping the legal landscape of anti-discrimination law, particularly in the context of emerging technologies such as AI.

#### **4.2.1 The EU Framework: A Comprehensive Legal Structure for Addressing Discrimination**

The European Union (EU) has long been at the forefront of promoting equality and combatting discrimination through its legislative framework, which is grounded in the principles of human dignity, equality, and respect for human rights. The EU's approach to anti-discrimination is notable for its binding directives and regulations that mandate equal treatment across its member states, compelling them to incorporate these standards into domestic law. These directives form the backbone of the EU's commitment to ensuring non-discrimination and promoting inclusivity in all areas of society, including those increasingly influenced by AI-driven technologies.

**Racial Equality Directive (2000/43/EC)**

Adopted in 2000, the Racial Equality Directive (2000/43/EC) is one of the cornerstones of EU anti-discrimination law. It specifically prohibits direct and indirect discrimination based on racial or ethnic origin across a broad spectrum of public and private life, including employment, education, healthcare, and access to goods and services.[[30]](#footnote-30) This directive obliges member states to establish effective remedies and enforcement mechanisms for individuals who experience discrimination.

One of the directive’s most critical provisions is the reversal of the burden of proof in discrimination cases: once a claimant has established facts from which it can be presumed that there has been discrimination, it falls on the defendant to prove that there was no breach of the principle of equal treatment. This procedural safeguard is particularly relevant in cases where AI systems are suspected of discriminatory outcomes, as it compels organizations to justify the fairness of their algorithms.[[31]](#footnote-31)

**Employment Equality Directive (2000/78/EC)**

The Employment Equality Directive (2000/78/EC) extends the EU’s anti-discrimination framework to employment and occupation, addressing discrimination based on religion or belief, disability, age, and sexual orientation. This directive is particularly significant in the context of AI-driven decision-making systems used in recruitment, promotions, and workplace management.

The directive requires member states to implement legislation that prohibits both direct and indirect discrimination in the workplace. It also promotes reasonable accommodation for individuals with disabilities, ensuring that AI systems used in recruitment or other employment contexts do not inadvertently disadvantage applicants with disabilities by failing to account for their specific needs.[[32]](#footnote-32)

The directive also underscores the need for proportionality in employment practices. AI systems used for hiring, for example, must not apply blanket exclusions or disproportionately Favor certain groups unless there is a legitimate objective, such as ensuring workplace safety, that justifies such measures. This principle is vital in guiding the development of fair and non-discriminatory AI systems in employment contexts.

**General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679)**

While the General Data Protection Regulation (GDPR) primarily addresses issues of data privacy and protection, it plays a pivotal role in mitigating discriminatory outcomes from AI-driven systems, particularly those that involve automated decision-making processes. AI systems are frequently powered by large datasets, and when such systems make decisions that significantly impact individuals—such as in credit scoring, recruitment, or law enforcement—their compliance with the GDPR becomes crucial.

Article 22 of the GDPR specifically addresses the rights of individuals concerning automated decision-making, including profiling. It grants individuals the right not to be subject to decisions made solely by automated processing, particularly when these decisions have legal or similarly significant effects, unless explicit consent is given, or certain contractual conditions are met. This article is particularly relevant in scenarios where AI systems are used to make decisions that might otherwise perpetuate bias, as it ensures that individuals can request human intervention and demand explanations for decisions made by AI systems.[[33]](#footnote-33)

Additionally, Article 9 of the GDPR places restrictions on the processing of sensitive personal data, such as racial or ethnic origin, political opinions, religious beliefs, or health information, unless specific exemptions apply. In the context of AI, this provision is critical because it prevents the use of discriminatory data in decision-making systems without a strong legal justification, reducing the risk of algorithmic bias and ensuring data protection and fairness.[[34]](#footnote-34) The GDPR's transparency requirements also provide a legal mechanism for individuals to challenge biased AI systems, as organizations must explain how personal data is processed and used in decision-making processes.

**The Proposed AI Act**

The EU is currently developing a groundbreaking legal framework specifically targeting AI technologies: the AI Act. Once adopted, the AI Act will represent the first comprehensive legal regime aimed at regulating AI across all member states. The AI Act distinguishes between AI systems based on their risk level and introduces stringent regulatory controls on those classified as "high-risk" systems—such as those used in critical areas like law enforcement, employment, and healthcare.[[35]](#footnote-35)

Under the AI Act, high-risk AI systems will be subject to mandatory requirements for transparency, accountability, and risk management. These systems will need to undergo extensive testing, certification, and ongoing monitoring to ensure they operate in compliance with EU principles of fairness and non-discrimination. The AI Act also emphasizes the importance of ensuring AI systems do not adversely affect fundamental rights, including the rights to privacy, equality, and non-discrimination.

One of the significant innovations of the AI Act is its focus on the ex-ante (before-the-fact) regulation of AI systems, requiring developers and deployers to anticipate and mitigate risks of bias before these systems are deployed. This preventive approach contrasts with traditional regulatory mechanisms, which often deal with discrimination ex-post (after-the-fact), offering the potential for more robust protection against algorithmic bias.[[36]](#footnote-36) Additionally, the AI Act introduces the concept of human oversight, which mandates that high-risk AI systems used in sensitive areas must always allow for human intervention or override to prevent harmful outcomes.

#### **4.2.2 The UK Framework: Anti-Discrimination Law in the Post-Brexit Era**

In the United Kingdom, the legal framework addressing discrimination has evolved considerably over the past decades. However, since Brexit, this framework has been shaped by the UK’s decision to retain many of the EU-derived anti-discrimination laws while simultaneously exploring ways to create more tailored legislation. The UK's approach to anti-discrimination post-Brexit continues to be primarily governed by the Equality Act 2010 and the Data Protection Act 2018, with the potential for future divergence from EU laws as the UK develops its own regulatory frameworks in areas such as Artificial Intelligence (AI).

**The Equality Act 2010**

The Equality Act 2010 is a landmark piece of legislation that consolidates and simplifies a wide range of pre-existing anti-discrimination laws into a single comprehensive Act. It provides robust protection against both direct and indirect discrimination, as well as harassment and victimization across various sectors, including employment, education, and access to goods and services. The Act covers nine protected characteristics: age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation.[[37]](#footnote-37)

The significance of the Equality Act in the context of AI-driven decision-making lies in its prohibition of discriminatory practices across these protected characteristics. The legislation makes it unlawful for organizations to implement processes, including those driven by AI, that disproportionately affect individuals based on these characteristics. This has become particularly important in sectors like recruitment, where AI tools are increasingly used to automate candidate screening. If these AI systems replicate existing biases from the training data, they may lead to outcomes that contravene the Equality Act. For instance, an AI hiring tool that disproportionately excludes women or candidates of a particular ethnic background could result in claims of both direct and indirect discrimination under the Act.[[38]](#footnote-38)

The Equality Act also provides mechanisms for individuals to challenge discriminatory practices, including the right to bring legal claims to employment tribunals or civil courts. This allows individuals who have been adversely impacted by biased AI systems to seek redress, providing an essential legal safeguard in a world where AI’s role in decision-making continues to grow.

**The Data Protection Act 2018**

The Data Protection Act 2018 (DPA 2018) incorporates the principles of the General Data Protection Regulation (GDPR) into UK law, ensuring that many of the GDPR’s provisions continued to apply after Brexit. Importantly, the DPA 2018 expands on GDPR provisions, tailoring data protection standards to the specific needs and context of the UK while maintaining continuity with EU data protection law.[[39]](#footnote-39)

The relevance of the DPA 2018 in the context of AI and anti-discrimination lies in its provisions on automated decision-making and the processing of sensitive personal data. Article 22 of the GDPR, incorporated into the DPA 2018, restricts automated decision-making that produces legal effects or significantly impacts individuals. This is particularly relevant for AI systems used in areas such as finance (credit scoring), recruitment, and law enforcement, where biased algorithms could lead to discriminatory outcomes. Under the DPA 2018, individuals have the right to challenge AI decisions, request human intervention, and receive an explanation for the decision-making process if it significantly impacts them.

Moreover, Section 10 of the DPA 2018 goes further by regulating the use of "special category data," including race, ethnicity, religion, and health data. This provision is crucial in mitigating the risk of bias in AI systems that rely on sensitive personal data. For example, if an AI system used by law enforcement disproportionately targets individuals of certain racial or ethnic backgrounds, it may violate both the DPA 2018 and the Equality Act 2010.[[40]](#footnote-40)

**Post-Brexit Considerations**

Following Brexit, the UK’s legal landscape has retained much of the EU-derived anti-discrimination law, particularly through the Equality Act 2010 and the DPA 2018, which incorporates GDPR principles. However, Brexit has created a distinct possibility that the UK will diverge from EU law in future regulatory developments, especially in the field of AI. This potential divergence could have significant implications for how AI-related discrimination is addressed in the UK compared to the EU.

The UK government has expressed interest in developing its own regulatory frameworks for AI, with an emphasis on innovation and flexibility. The government’s National AI Strategy, published in 2021, highlights the need to balance innovation with public trust, setting the stage for future AI-specific legislation.[[41]](#footnote-41) The UK has also introduced the Data Protection and Digital Information Bill (DPDI Bill) in 2022, aimed at reforming data protection laws post-Brexit. This bill seeks to amend the DPA 2018, reduce burdens on businesses, and introduce new provisions for AI and automated decision-making systems.[[42]](#footnote-42)

While the DPDI Bill and future AI regulations may provide greater clarity and adaptability for businesses, there are concerns that such divergence could weaken protections against AI-driven discrimination. If the UK were to reduce the obligations on organizations to explain and justify AI-driven decisions, individuals might face greater challenges in contesting biased outcomes. Additionally, as the EU moves forward with the AI Act, which imposes strict rules on high-risk AI systems, any deviation by the UK could lead to regulatory discrepancies that impact the protection of individuals from discrimination in AI-driven contexts.

The UK’s judicial system will also play a key role in shaping how anti-discrimination laws are applied to AI-driven decision-making systems. UK courts have already begun to engage with the complexities of automated decision-making, as demonstrated in cases like *R (on the application of Bridges) v Chief Constable of South Wales Police*, where the Court of Appeal found that the use of facial recognition technology by the police was unlawful because it lacked adequate safeguards to prevent discrimination.[[43]](#footnote-43)

As AI continues to evolve, UK courts will likely see more cases challenging the use of AI in areas such as recruitment, financial services, and law enforcement, particularly where it results in discriminatory outcomes. These legal challenges will be essential in determining how the UK’s anti-discrimination framework adapts to the unique challenges posed by AI technologies.

**Conclusion**

The UK’s anti-discrimination framework, anchored by the Equality Act 2010 and the Data Protection Act 2018, provides strong protections against discriminatory practices, including those driven by AI systems. However, the post-Brexit legal landscape presents both opportunities and risks

**4.3 Case Law Analysis**

The application of anti-discrimination laws to AI systems is a relatively new but rapidly evolving area of law. As AI becomes more deeply integrated into decision-making processes, courts across the European Union (EU) and the United Kingdom (UK) are increasingly called upon to interpret existing legal frameworks considering AI-driven technologies. These cases highlight the challenges and opportunities of applying anti-discrimination and data protection laws to AI systems, revealing gaps in current regulations and underscoring the need for legal reforms. The following sections explore key cases in the EU and UK that offer insights into how courts and regulatory bodies are navigating this emerging legal terrain.

#### **4.3.1 EU Case Law**

The European Union has been at the forefront of regulating AI systems, with data protection laws such as the General Data Protection Regulation (GDPR) and the European Convention on Human Rights (ECHR) serving as pivotal legal frameworks. The following cases demonstrate how these laws have been applied to AI and ADM systems, particularly in addressing concerns related to transparency, accountability, and discrimination.

**SyRI Welfare Automation in the Netherlands (2020) (Case number: *Rechtbank Den Haag, 5 februari 2020, ECLI:NL: RBDHA:2020:865)***  
In the Netherlands, the SyRI (System Risk Indication) case is a key example of AI-driven automated decision-making (ADM) systems resulting in discriminatory practices. SyRI was developed by the Dutch government to detect welfare fraud by profiling individuals using large datasets, including income and household data. Civil rights groups challenged SyRI, arguing it disproportionately targeted low-income and immigrant communities, raising concerns about privacy violations and indirect discrimination.

In 2020, the District Court of The Hague ruled that SyRI violated the European Convention on Human Rights (ECHR), particularly Article 8 (right to privacy) and Article 14 (prohibition of discrimination). The court found that the system lacked transparency and safeguards, leading to disproportionate targeting of vulnerable groups, thus constituting indirect discrimination[[44]](#footnote-44) This ruling set a critical precedent, reinforcing the need for ADM systems to comply with both GDPR and human rights standards. It underscored the importance of transparency and proportionality in preventing discriminatory outcomes, especially when ADM systems affect marginalized groups.

**4.3.2 UK Case Law**

In the UK, courts have similarly been confronted with legal challenges involving AI systems, particularly where automated decision-making has led to discriminatory outcomes. The Equality Act 2010 and Data Protection Act 2018 provide the primary legal frameworks through which ADM systems are scrutinized, especially when issues of privacy and indirect discrimination arise.

**R (on the application of Bridges) v Chief Constable of South Wales Police (2020) (Case number: *R (on the application of Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058, [2020] 1 WLR 672):**  
The Bridges case involved the use of Automated Facial Recognition (AFR) technology by South Wales Police, which Edward Bridges argued violated his rights under the Data Protection Act 2018 and Human Rights Act 1998, citing privacy concerns and the risk of racial discrimination. Bridges contended that AFR disproportionately misidentified individuals from minority ethnic backgrounds, a well-documented issue in facial recognition technology[[45]](#footnote-45).

In 2020, the Court of Appeal ruled that the use of AFR was unlawful, finding it failed to comply with data protection and human rights laws, particularly regarding privacy and discrimination. The court emphasized the lack of safeguards to prevent misuse, especially the risk of misidentification of marginalized groups[[46]](#footnote-46). This ruling set a key precedent in the UK, establishing clear standards for the use of AI in law enforcement and reinforcing the need for transparency and accountability, particularly where racial bias is a concern.

**R (on the application of Edward Santon) v Secretary of State for Work and Pensions (2020) (Case number: *R (on the application of Edward Santon) v Secretary of State for Work and Pensions* [2020] EWHC 381 (Admin), [2020] PTSR 1562)**  
In the Santon case, Edward Santon, a claimant with disabilities, challenged the use of automated decision-making by the Department for Work and Pensions (DWP), arguing it disproportionately disadvantaged disabled individuals by failing to consider their specific needs. The High Court ruled in 2020 that the system led to indirect discrimination by lacking transparency and human oversight, violating both the Equality Act 2010 and Article 8 of the Human Rights Act 1998[[47]](#footnote-47). This case underscores the limitations of automated decision-making (ADM) systems in public services, highlighting the need for human oversight and transparency to prevent discrimination and ensure fairness

### **4.4 Conclusion**

The anti-discrimination laws in the EU and UK provide a strong foundation for addressing discrimination in various contexts, including AI-driven decision-making. However, the unique challenges posed by AI such as the opacity of algorithms, the complexity of AI systems, and the potential for exacerbating existing biases highlight significant gaps in these legal frameworks. In Chapter 5, this thesis evaluates the effectiveness of current legal frameworks in the UK and EU.

## **Chapter 5: Evaluating the Effectiveness of Current Legal Frameworks**

### **5.1 Introduction**

This chapter critically evaluates the effectiveness of current legal frameworks in the EU and UK in addressing and mitigating biases in AI-driven decision-making processes. It assesses the strengths and weaknesses of these frameworks and considers their adequacy in protecting individuals from discrimination resulting from automated systems.

### **5.2 Strengths of Existing Legal Frameworks**

The legal frameworks in both the EU and UK offer several significant strengths in addressing discrimination, particularly in the context of emerging technologies like artificial intelligence (AI). These frameworks are built on well-established anti-discrimination principles and data protection laws, which provide crucial protections for individuals and ensure a degree of accountability for organizations using AI in decision-making processes.

One of the major strengths of these frameworks is their comprehensive coverage of protected characteristics. In the EU, the Racial Equality Directive (2000/43/EC) and Employment Equality Directive (2000/78/EC) are key components that protect individuals from discrimination based on race, ethnicity, religion, disability, age, and sexual orientation.[[48]](#footnote-48) Similarly, in the UK, the Equality Act 2010 consolidates multiple anti-discrimination laws into a single legislative framework, ensuring broad protection for individuals across various contexts, including employment, education, and access to goods and services. This comprehensive approach ensures that the law can address the potential for discrimination in AI-driven systems, which often process vast amounts of personal data and can have significant impacts on people's lives.

Another notable strength of these frameworks is the recognition of indirect discrimination, particularly important in the context of AI systems. Indirect discrimination occurs when a seemingly neutral rule or practice disproportionately impacts a particular group. Given that many AI systems are trained on historical data that may reflect societal biases, they can unintentionally replicate or even amplify these biases in their decision-making processes. The recognition of indirect discrimination in the Equality Act 2010 and EU directives is critical in addressing these biases, as it allows for legal challenges even when the discriminatory effect is not intentional but arises from the underlying data or algorithmic processes. For example, AI systems used in recruitment may disproportionately disadvantage women or minority candidates if trained on biased historical data, even if the system itself does not explicitly consider gender or race.[[49]](#footnote-49)

Additionally, the General Data Protection Regulation (GDPR) in the EU provides robust protections against discriminatory data processing. The GDPR’s provisions on automated decision-making and profiling, especially Article 22, grant individuals the right not to be subject to decisions based solely on automated processing that produces significant legal or similarly consequential effects. This is particularly important in the AI context, where decisions related to credit scoring, employment, or even criminal justice can have life-altering consequences for individuals.[[50]](#footnote-50) Moreover, the GDPR mandates transparency in data processing, requiring organizations to inform individuals about the logic behind automated decisions (Articles 13 and 14). This transparency is crucial for identifying and challenging biased AI systems, giving individuals the tools they need to understand how decisions about them are being made and whether those decisions may be rooted in discriminatory practices.

Furthermore, the protection of sensitive data under the GDPR is another strength in mitigating AI-related discrimination. Article 9 of the GDPR prohibits the processing of special categories of personal data, such as race, ethnicity, and political opinions, unless specific conditions are met.[[51]](#footnote-51) This is a vital safeguard in AI-driven systems, as it ensures that sensitive personal data cannot be used in a way that could lead to discriminatory outcomes. In high-stakes areas such as healthcare, law enforcement, and finance, these protections are especially crucial, as the misuse of sensitive data in AI systems could disproportionately affect marginalized groups.

Judicial interpretation has also played a key role in reinforcing the strengths of these legal frameworks. Courts in both the EU and UK have shown a willingness to adapt existing legal principles to address the challenges posed by AI. In cases such as R (on the application of Bridges) v Chief Constable of South Wales Police, the UK courts scrutinized the use of automated facial recognition technology, applying data protection and anti-discrimination laws to ensure that its deployment did not infringe on individuals' rights.[[52]](#footnote-52) This case demonstrates the capacity of the judiciary to interpret existing laws in ways that address the complexities of AI, providing an additional layer of accountability for organizations using these technologies.

However, while these frameworks provide robust protections, their effectiveness ultimately depends on enforcement and implementation. The courts and regulatory bodies need to continue playing a proactive role in ensuring that AI systems comply with legal standards, particularly in contexts where the risks of discrimination are high. The potential for indirect discrimination and the challenges associated with algorithmic opacity require ongoing vigilance from both the judiciary and regulators to ensure that AI technologies are used responsibly and fairly.

### **5.3 Weaknesses and Gaps in the Legal Frameworks**

Despite the strengths outlined above, there are significant weaknesses and gaps in the current legal frameworks that limit their effectiveness in addressing AI-related discrimination. These weaknesses stem from the inherent complexities of AI systems and the difficulties in applying traditional legal concepts to modern technological contexts.

#### **5.3.1 Opacity and the “Black Box” Problem**

One of the most significant challenges in regulating AI is the opacity of AI systems, often referred to as the “black box” problem. Many AI algorithms, particularly those based on machine learning and neural networks, operate in ways that are not easily interpretable by humans. This lack of transparency makes it difficult to understand how decisions are made, let alone identify and prove discrimination.[[53]](#footnote-53)

Current legal frameworks, including the GDPR, emphasize transparency and the right to an explanation, but these provisions are often insufficient in practice. The requirement to explain the logic behind automated decision-making processes, as stipulated in Articles 13-15 of the GDPR, is challenging to fulfil for complex AI systems where even developers may not fully understand the decision-making process.[[54]](#footnote-54) This gap between legal requirements and technological capabilities undermines the effectiveness of anti-discrimination laws in the context of AI, as it limits the ability of individuals to challenge biased decisions.

#### **5.3.2 Limitations of the GDPR in Addressing AI Discrimination**

While the General Data Protection Regulation (GDPR) is lauded for establishing a comprehensive framework for data protection, its ability to fully address AI-related discrimination remains limited. Article 22, which addresses automated decision-making, has been criticized for its ambiguity and narrow scope. The provision applies only to decisions “based solely” on automated processing, excluding many AI systems that involve some level of human oversight or intervention[[55]](#footnote-55). In practice, many AI systems are designed with minimal human involvement often just a cursory review which might technically exempt them from the protections offered by Article 22. This raises the risk that organizations can use nominal human intervention to bypass the restrictions of fully automated decision-making, even when the system is functionally AI-driven and prone to discriminatory outcomes.

For example, in recruitment processes, AI systems may screen applicants with minimal human input, meaning the decision-making process is not "solely" automated under the strict interpretation of the GDPR. As a result, companies may claim compliance with GDPR while still exposing individuals to biased, AI-driven decisions. This loophole is a significant limitation because it ignores the reality that even limited human involvement does not necessarily mitigate AI biases particularly when human reviewers tend to trust AI outputs without scrutinizing them critically. The potential for algorithmic bias, especially in high-stakes areas like hiring, lending, and criminal justice, remains a pressing concern. The narrow interpretation of "solely" automated decisions fails to address these real-world complexities, thereby diminishing the GDPR's effectiveness in curbing AI-related discrimination.

Moreover, the GDPR’s emphasis on consent as a legal basis for processing personal data can be problematic in the context of AI. Consent is often not meaningful when individuals lack a genuine understanding of how their data will be used or the potential risks of automated decision-making.[[56]](#footnote-56) Individuals are frequently asked to provide consent without fully understanding how their data will be processed or how automated systems work. The complexity and opacity of AI systems make it extremely difficult for the average individual to comprehend the long-term implications of data use. For example, in predictive policing or credit scoring, individuals may consent to their data being used without realizing that this data could feed into an AI system that perpetuates existing biases or leads to unfair treatment based on historical inequalities.

Moreover, the concept of consent assumes a level of autonomy and informed choice that is often absent in real-world scenarios. Organizations may present consent requests in ways that are overwhelming or misleading, particularly through dense and jargon-heavy privacy policies. In practice, individuals often have little real power to refuse consent, especially in situations where they depend on the service in question, such as social media platforms, financial institutions, or public services. This undermines the core principle of informed consent, as individuals may agree to terms without fully grasping the risks associated with automated decision-making, including potential discriminatory outcomes. In these contexts, consent is more of a formality than a genuine protective mechanism, making it an inadequate safeguard against AI bias.

Furthermore, the GDPR’s focus on data protection and privacy does not necessarily translate into fairness in decision-making. The regulation is designed to protect personal data, but it does not specifically address the ways in which AI systems might process this data in ways that produce discriminatory outcomes. While the GDPR does contain provisions that prevent the use of sensitive data (such as race, ethnicity, or political beliefs) without explicit consent, AI systems can still inadvertently use proxy variables such as ZIP codes or purchasing behavior that indirectly correlate with these protected characteristics, resulting in discriminatory effects. As a result, the GDPR lacks the necessary mechanisms to detect and prevent such indirect discrimination, which limits its ability to combat AI bias comprehensively.

**5.3.3 Insufficient AI-Specific Legislation**

One of the critical shortcomings of the current legal frameworks in both the EU and UK is the lack of AI-specific legislation that adequately addresses the unique challenges posed by AI technologies. While the GDPR and Equality Act 2010 offer some mechanisms to combat AI-related discrimination, these regulations were not originally designed with AI in mind. As a result, they fall short of addressing the full scope of issues that arise from the increasing use of AI systems in decision-making processes. The regulatory gaps are particularly evident in the areas of algorithmic accountability, bias detection, and oversight elements that are fundamental to ensuring that AI technologies do not perpetuate or exacerbate existing inequalities.

The GDPR, for example, focuses primarily on data protection and privacy but does not provide sufficient guidance on the ethical and social implications of AI systems that process large amounts of data for decision-making. Similarly, while the Equality Act 2010 addresses direct and indirect discrimination based on protected characteristics, it was created in a pre-AI era, meaning it lacks specific provisions that account for the complexities of algorithmic bias. For instance, AI systems can inadvertently generate discriminatory outcomes based on proxy variables like postal codes or educational backgrounds, which, while not explicitly protected characteristics, can correlate with race, socioeconomic status, or other protected attributes. The current legal frameworks do not fully address these indirect pathways of bias, leaving significant regulatory blind spots.

The EU’s proposed AI Act marks a significant development in addressing these gaps, as it represents an attempt to create a comprehensive legal framework specifically tailored to AI. The AI Act aims to establish harmonized rules for the development, deployment, and use of AI systems across the EU, focusing heavily on risk management and the prevention of discriminatory outcomes. By classifying AI systems based on their risk levels ranging from minimal to unacceptable the AI Act seeks to ensure that high-risk AI systems, such as those used in critical areas are subject to strict scrutiny and compliance requirements. These include provisions for algorithmic transparency, human oversight, and robust auditing processes to identify and mitigate bias[[57]](#footnote-57).

However, despite the AI Act’s promise, it is still in the proposal stage and has yet to be implemented. Moreover, its success will depend not only on its adoption but also on how effectively it is enforced across the diverse legal and regulatory landscapes of EU member states. The delay in implementing AI-specific legislation leaves a regulatory vacuum, allowing AI technologies to proliferate without sufficient legal safeguards in place. This gap is particularly concerning given the rapid advancement of AI and its widespread integration into both public and private sectors. Until the AI Act is finalized and enforced, organizations may continue to exploit loopholes in existing legislation, potentially perpetuating discriminatory outcomes.

The UK, in contrast, faces an even more precarious situation, particularly in the post-Brexit landscape. The absence of AI-specific legislation in the UK leaves a considerable gap in its regulatory framework. While the UK continues to rely on the GDPR and the Equality Act 2010 as primary tools for addressing issues of discrimination, these laws are insufficient for tackling the nuanced challenges presented by AI. Brexit has complicated the situation further, as the UK is no longer obligated to follow the EU’s evolving legal standards, including the forthcoming AI Act. This divergence creates regulatory uncertainty and could potentially weaken protections against AI bias, as there is no clear plan for the UK to develop AI-specific regulations that match or exceed those being proposed in the EU[[58]](#footnote-58).

Moreover, without AI-specific laws, the UK's reliance on existing anti-discrimination frameworks becomes increasingly problematic. As AI systems become more complex and autonomous, the ability of courts to apply traditional legal principles to these new technologies is limited. For example, while the Equality Act prohibits indirect discrimination, the complexities of algorithmic decision-making may make it difficult for courts to determine causality or intent behind biased outcomes. AI systems often operate as black boxes; therefore, this opacity creates significant challenges in proving that discrimination has occurred and in holding organizations accountable for AI-driven decisions.

The absence of AI-specific legislation also raises concerns about ethical governance and the broader societal impact of AI. Without clear legal guidelines that address algorithmic fairness and accountability, there is a risk that organizations will prioritize efficiency and profit over social responsibility. The lack of regulation means that the burden of identifying and addressing AI bias often falls on individuals who may lack the resources or knowledge to challenge automated decisions. This creates a power imbalance where the most vulnerable individuals those disproportionately affected by biased algorithms are least equipped to defend their rights.

**5.3.4 Enforcement Challenges and Resource Constraints**

Even where legal frameworks are in place, enforcement poses a significant challenge. Regulatory bodies tasked with overseeing compliance with anti-discrimination and data protection laws often lack the technical expertise and resources needed to effectively monitor and regulate AI systems. This is particularly problematic given the rapid pace of technological advancement in AI.[[59]](#footnote-59)

Moreover, individuals who are subject to discriminatory AI decisions may face significant barriers in accessing justice. The complexity of AI systems, coupled with the costs and time involved in pursuing legal action, can discourage individuals from challenging biased outcomes. This creates a situation where discriminatory practices may persist simply because they go unchallenged.[[60]](#footnote-60)

### **5.5 Conclusion**

The current legal frameworks in the EU and UK offer a strong foundation for protecting individuals from discrimination, but they are not fully equipped to address the unique challenges posed by AI systems. The opacity of AI algorithms, the limitations of existing laws like the GDPR, the absence of AI-specific legislation, and enforcement challenges all contribute to gaps in the legal protections against AI-related discrimination.

To address these challenges, it is essential to develop a more comprehensive and nuanced legal framework that is specifically tailored to the complexities of AI. This framework should include provisions for algorithmic transparency, regular audits, and robust accountability mechanisms. Additionally, the EU and UK must invest in the resources and expertise needed to enforce these laws effectively. By strengthening the legal framework for AI, both jurisdictions can better protect individuals from discrimination and ensure that AI technologies are used in a manner that upholds principles of fairness, equality, and human rights.

## **Chapter 6: Legal Reforms and Future Directions**

### **6.1 Introduction**

The growing integration of Artificial Intelligence (AI) in decision-making processes across various sectors necessitates a re-evaluation of current legal frameworks. While the existing anti-discrimination laws in the European Union (EU) and the United Kingdom (UK) provide a foundation for addressing biases, they are not fully equipped to contend with the complexities of AI technologies. This chapter explores the need for legal reforms to enhance the effectiveness of these frameworks and proposes future directions to better address the challenges posed by AI. The chapter also considers the role of international cooperation and the importance of developing a human rights-centric approach to AI governance.

### **6.2 Enhancing AI Transparency and Accountability**

#### **6.2.1 Legal Mandates for Explainability**

One of the critical challenges in regulating AI is the opacity of AI decision-making processes, often referred to as the "black box" problem.[[61]](#footnote-61) This lack of transparency complicates the assessment of how decisions are made, especially when these decisions result in discriminatory outcomes. To address this issue, legal reforms should include explicit mandates for explainability, ensuring that AI systems provide clear, understandable explanations of their decision-making processes.

The General Data Protection Regulation (GDPR) in the EU already includes provisions requiring data controllers to inform individuals about the logic involved in automated decision-making processes.[[62]](#footnote-62) However, these provisions need strengthening to ensure AI systems generate explanations that are both technically accurate and comprehensible to non-experts. Achieving this requires developing standardized practices for generating explanations and integrating these practices into AI system design and deployment from the outset.

Implementing these reforms will necessitate overcoming several barriers:

1. **Political Challenges**: Lawmakers may face resistance from industry stakeholders who argue that extensive explainability requirements could stifle innovation or reveal proprietary algorithms. Effective advocacy and stakeholder engagement will be crucial to address these concerns and build broad support for reform.
2. **Technological Challenges**: Developing standardized methods for explainability that are both accurate and understandable presents technical difficulties. Collaboration between technologists and regulators will be essential to create practical, achievable standards.
3. **Institutional Challenges**: Regulators and auditing bodies may need to build new capabilities and frameworks to assess AI systems effectively. This may involve training personnel, developing new tools, and establishing procedures for conducting comprehensive audits.

In addition to enhancing transparency, there is a pressing need for robust accountability mechanisms to ensure that AI systems do not perpetuate or exacerbate discrimination. Algorithmic accountability can be achieved through regular audits of AI systems, which should be mandated by law for high-risk applications.[[63]](#footnote-63) These audits would involve a comprehensive evaluation of the AI system's design, data inputs, and decision-making processes to identify and mitigate biases.

Algorithmic audits could involve the systematic examination of AI systems to assess their compliance with ethical standards and legal requirements, including anti-discrimination laws. These audits can identify biases in AI models, evaluate the fairness of decision-making processes, and recommend corrective actions where necessary.[[64]](#footnote-64)

However, the effectiveness of algorithmic audits will depend on several factors, including the transparency of the AI system, the quality of the audit process, and the willingness of organizations to implement audit recommendations. To enhance the role of algorithmic audits, it will be necessary to establish legal requirements for regular audits, particularly for high-risk AI systems, as proposed in the EU’s AI Act.[[65]](#footnote-65) The proposed AI Act in the EU is a step in this direction, as it requires high-risk AI systems to undergo conformity assessments, including evaluations of their impact on fundamental rights.[[66]](#footnote-66) However, similar provisions need to be adopted and rigorously enforced in the UK, particularly considering its departure from the EU.

In addition to algorithmic audits, robust accountability mechanisms are essential for addressing AI-related discrimination. These mechanisms include clear lines of responsibility for AI decisions, transparency obligations, and the ability to hold organizations accountable for biased outcomes.

One approach to enhancing accountability is the implementation of a “right to contest” AI decisions, allowing individuals to challenge decisions that have been made by or with the assistance of AI. This right is supported by the GDPR but could be strengthened through additional legal provisions that ensure individuals have meaningful opportunities to contest and rectify discriminatory decisions.[[67]](#footnote-67)

Furthermore, accountability can be reinforced through the development of industry standards and best practices for AI development and deployment. These standards can provide guidance on issues such as data quality, algorithmic fairness, and transparency, helping organizations to minimize the risk of discrimination and ensuring that AI systems are used responsibly.

### **6.3 Strengthening Data Protection and Bias Mitigation**

Bias in AI systems often stem from the data used to train these models. When training data is unrepresentative of the population or reflects historical biases, the resulting AI systems are likely to produce biased outcomes.[[68]](#footnote-68) Addressing this issue requires legal reforms that go beyond current frameworks, particularly in mandating the use of diverse and representative datasets. Such reforms should include robust bias impact assessments during data collection and processing to identify and mitigate potential risks early. Plus, the introduction of data provenance requirements would ensure transparency regarding the origin, quality, and history of the data used in AI, facilitating regulatory oversight and accountability.

Additionally, the concept of “data provenance” should be integrated into legal frameworks, ensuring that the origin, quality, and processing history of data used in AI systems are documented and transparent. This would enable auditors and regulators to trace the sources of bias and hold organizations accountable for the quality of the data they use.[[69]](#footnote-69)

On the other hand, Prohibiting the Use of Certain Sensitive Datacan also play a crucial role, While the GDPR prohibits the processing of special categories of personal data, such as race, ethnicity, and religion, unless specific conditions are met, it has been evident that these provisions should be expanded to address the unique risks associated with AI[[70]](#footnote-70). Therefore, Legal reforms should include stricter regulations on the use of sensitive data in AI systems, particularly in high-stakes areas such as law enforcement, healthcare, and financial services.

Moreover, exceptions that allow for the processing of sensitive data should be narrowly defined and subject to rigorous oversight. This includes requiring organizations to demonstrate that the use of such data is necessary, proportionate, and accompanied by adequate safeguards to prevent discrimination[[71]](#footnote-71).

By tightening restrictions and enhancing accountability, legal frameworks can more effectively prevent AI systems from perpetuating and amplifying societal biases.

### **6.4 Developing AI-Specific Anti-Discrimination Legislation**

#### **6.4.1 The Need for AI-Specific Legal Instruments**

Current anti-discrimination laws were not designed with AI technologies in mind, leading to significant gaps in their applicability to AI-driven decision-making systems[[72]](#footnote-72). As AI continues to evolve and become more pervasive, there is an urgent need for AI-specific anti-discrimination legislation that addresses the unique challenges posed by these technologies.

Such legislation should include clear definitions of what constitutes discriminatory AI practices, along with provisions that explicitly prohibit these practices. It should also establish standards for fairness in AI systems, drawing on interdisciplinary research from fields such as computer science, ethics, and law.[[73]](#footnote-73) By creating a dedicated legal framework for AI, policymakers can ensure that anti-discrimination principles are fully integrated into the design, deployment, and oversight of AI systems.

Additionally, to ensure that AI technologies are developed and used in a manner that respects human rights, AI-specific legislation should be grounded in a human rights-centric approach to governance. This approach would prioritize the protection of individual rights and freedoms, including the right to non-discrimination, over the commercial or operational interests of AI developers and users.

A human rights-centric approach would also require legal frameworks to incorporate principles such as equality, fairness, and accountability into AI governance.[[74]](#footnote-74) This could be achieved through the establishment of independent regulatory bodies with the authority to oversee AI systems, investigate complaints, and enforce compliance with anti-discrimination laws. These bodies should be empowered to impose penalties for non-compliance and require remedial actions where AI systems are found to have violated individuals’ rights.

### **6.5 International Collaboration and Harmonization of Standards**

AI is a global technology, and its impact transcends national borders. As such, effective regulation of AI-related discrimination requires international collaboration and the harmonization of legal standards across jurisdictions.[[75]](#footnote-75) This is particularly important given the cross-border nature of many AI systems, which are often developed in one country, deployed in another, and used by individuals from various regions.

International collaboration can help to ensure that AI systems are subject to consistent legal standards, regardless of where they are developed or used. It can also facilitate the sharing of best practices, promote the development of global norms for AI governance, and prevent regulatory arbitrage, where organizations exploit differences in national regulations to evade accountability[[76]](#footnote-76).

To achieve harmonization of legal standards, international organizations such as the United Nations (UN), the Organisation for Economic Co-operation and Development (OECD), and the Council of Europe should play a leading role in developing global standards for AI governance. These standards should address key issues such as algorithmic transparency, data protection, and non-discrimination, and should be aligned with international human rights law.[[77]](#footnote-77)

The development of international standards should involve a multi-stakeholder approach, bringing together governments, industry, civil society, and academia to ensure that the standards are comprehensive, inclusive, and reflective of diverse perspectives. Additionally, mechanisms for monitoring and enforcing compliance with these standards should be established, to ensure that they are effectively implemented and upheld by all stakeholders.[[78]](#footnote-78)

### **6.6 Conclusion**

The rapid advancement of AI technologies presents significant challenges to existing legal frameworks, particularly in anti-discrimination law. While the EU and UK have made strides in addressing some of these challenges, there is a clear need for legal reforms that specifically target the unique risks posed by AI. Enhancing transparency and accountability, strengthening data protection, developing AI-specific anti-discrimination legislation, and promoting international collaboration are critical steps toward ensuring that AI technologies are developed and used in a manner that upholds principles of fairness, equality, and human rights.

Moving forward, it is essential that policymakers, regulators, and other stakeholders work together to create a comprehensive and coherent legal framework for AI. By doing so, they can help to mitigate the risks of discrimination, protect individuals’ rights, and ensure that AI technologies contribute positively to society.

## **Chapter 7: Conclusion**

### **7.1 Summary of Findings**

This This thesis critically assesses the effectiveness of anti-discrimination laws in the EU and UK regarding biases in AI automated decision-making systems. AI's role in sectors such as finance, healthcare, and law enforcement underscore its significant impact on individuals and communities. As AI systems become more prevalent, they introduce new challenges to existing legal frameworks designed to prevent discrimination.

Chapter 3 examined how biases can be embedded in AI systems through data, algorithms, and contextual factors, illustrating the potential for AI to perpetuate societal inequalities. Chapter 4 reviewed key EU and UK anti-discrimination laws, including the Racial Equality Directive, GDPR, and the Equality Act 2010. It identified significant gaps in these frameworks, such as the opacity of AI decision-making processes and the lack of AI-specific legislation.

Chapter 5 analysed the strengths and weaknesses of current legal frameworks, noting their comprehensive coverage but also highlighting issues like the “black box” problem, limitations of GDPR Article 22, and enforcement challenges. Chapter 6 proposed reforms to address these challenges, advocating for enhanced AI transparency, stronger data protection measures, AI-specific anti-discrimination legislation, and international collaboration to harmonize AI governance standards.

**7.2 Implications for Policy and Practice**

The findings of this thesis highlight critical implications for both policy and practice in the regulation of AI technologies. Policymakers in the EU and UK must prioritize the creation of AI-specific legislation that confronts the unique challenges posed by AI systems.

Existing legal frameworks, while robust in some areas, are insufficient to address the complexities of AI-driven decision-making. Legislation should focus on clearly defining and prohibiting discriminatory AI practices, establishing standards for fairness, and ensuring these standards are consistently enforced through stringent accountability mechanisms.

Furthermore, the notion of AI transparency needs to be reimagined. Transparency should extend beyond technical explainability and must focus on making AI decisions both understandable and contestable by individuals affected by these systems. This necessitates integrating explainability into the design phase of AI development, as well as creating standardized methods for generating and communicating explanations in a way that is accessible to the general public. Without these reforms, AI risks becoming an opaque tool that entrenches rather than mitigates inequality.

In practice, organizations using AI systems must adopt proactive measures to audit and review their algorithms regularly, identifying and addressing any biases uncovered through these processes. This requires not only technical proficiency but also a strong commitment to ethical AI deployment.

The thesis underscores the need for companies to be legally accountable for the consequences of their AI systems, with meaningful sanctions imposed for failing to meet anti-discrimination standards. Additionally, data protection laws must be strengthened, particularly in terms of ensuring the use of diverse and representative datasets and enforcing stricter controls on the processing of sensitive personal data. Such measures are essential to mitigate bias and prevent AI from exacerbating existing societal inequalities.

On a broader scale, the thesis calls for international cooperation in AI governance, advocating for the alignment of AI regulations with international human rights principles. Given the cross-border nature of AI technologies, harmonized global standards are crucial to ensure that AI systems do not infringe upon the rights to equality and non-discrimination.

**7.3 Recommendations for Further Research**

While this thesis has provided a comprehensive analysis of the legal challenges posed by AI in the context of anti-discrimination law, it also opens avenues for further research. Future studies could explore the following areas:

1. **Comparative Analysis of AI-Specific Legislation**: As more jurisdictions develop AI-specific laws, a comparative analysis of these legal frameworks could provide valuable insights into best practices for regulating AI. Such research could examine how different countries address issues of algorithmic bias, transparency, and accountability, and what lessons can be learned from these approaches.
2. **Impact of AI on Intersectional Discrimination**: Further research is needed to explore how AI systems impact individuals who belong to multiple protected groups, such as those facing intersectional discrimination based on race, gender, and disability. Understanding these dynamics is crucial to developing legal frameworks that adequately protect all individuals from AI-driven discrimination.
3. **Effectiveness of Algorithmic Audits**: While algorithmic audits are increasingly recognized as a tool for mitigating AI bias, there is a need for empirical research to assess their effectiveness in practice. Future studies could investigate how audits are conducted, the challenges faced by auditors, and the extent to which audits lead to meaningful changes in AI systems.
4. **Public Perception and Trust in AI**: The success of legal reforms and AI governance depends in part on public perception and trust in AI technologies. Research into public attitudes towards AI, particularly in the context of fairness and discrimination, could inform the development of policies that address public concerns and build trust in AI systems.

### **7.4 Final Thoughts**

The integration of AI into decision-making processes signals both unprecedented opportunities and formidable challenges for modern societies. On one hand, AI holds the promise of revolutionizing industries, streamlining public services, and improving decision-making efficiency. Yet, on the other hand, it carries the inherent risk of perpetuating systemic biases, replicating historical inequalities, and amplifying discriminatory practices. The nuanced complexity of AI lies in its potential to both promote progress and entrench societal inequalities, depending on how it is designed, deployed, and regulated.

The existing legal frameworks in the EU and UK, such as the General Data Protection Regulation (GDPR) and the Equality Act 2010, provide valuable mechanisms for mitigating the risks associated with AI. However, these frameworks were not developed with the intricacies of AI in mind and, as such, struggle to fully address the unique challenges posed by automated decision-making systems. Plus, the opaque nature of many AI models, the "black box" problem, and the lack of meaningful accountability in certain applications demonstrate the gaps in current regulatory approaches.

Additionally, the right to explanation under the GDPR, while critical, has proven difficult to enforce meaningfully in complex AI contexts, leaving individuals vulnerable to opaque, and often biased, automated decisions. Similarly, the indirect discrimination provisions of the Equality Act, while helpful, struggle to keep pace with the scale and speed at which AI systems can impact large populations.

Therefore, Legal reforms are not only necessary but urgent. AI systems must be designed and governed in ways that are transparent, accountable, and fully aligned with the principles of fairness and equality. This includes embedding ethical considerations at every stage of AI development, from data collection to algorithmic design and deployment. Human oversight must be strengthened in areas where AI decisions have significant consequences on individuals’ lives, especially in high-risk sectors like criminal justice, finance, healthcare, and welfare. Furthermore, the role of algorithmic audits and impact assessments must be formalized and expanded to identify and mitigate potential biases before these systems are deployed at scale.

Beyond regulation, the discourse surrounding AI must evolve. Rather than merely reacting to technological advances, policymakers, academics, and civil society must engage proactively to shape the development of AI technologies. This includes fostering interdisciplinary collaboration between technologists, ethicists, lawyers, and policymakers to ensure that AI is not just technologically sound but also socially just. As a result, public trust in AI systems can be built not only on their technical capabilities but also on their perceived fairness, accountability, and alignment with fundamental human rights.

To conclude, the future of AI is not predetermined; it is a human-made construct that reflects the values and priorities we choose to embed in it. As AI continues to evolve, so too must the legal, regulatory, and ethical frameworks that govern its use. Only through sustained, collaborative efforts can we ensure that AI becomes a force for good, a technology that enhances, rather than undermines, equality, fairness, and human dignity.

Top of Form

Bottom of Form

## **References**

Top of Form

Bottom of Form

Top of Form

Bottom of Form

Abbott, R., (2020). The Reasonable Robot: Artificial Intelligence and the Law. Cambridge University Press.

AI Council, (2021). AI Roadmap. [pdf] Available at: <https://www.gov.uk/government/publications/ai-roadmap>

AI Ethics Lab, (2021). AI Ethics Lab: Bridging the Gap Between Research and Practice. Available at: <https://aiethicslab.com>

AI HLEG, (2019). Ethics Guidelines for Trustworthy AI. [pdf] Available at: <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419> [Accessed 12 June 2024].

Ananny, M. and Crawford, K., (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(3), pp.973-989.

Andre Esteva and others, 'A Guide to Deep Learning in Healthcare' (2019) 25 *Nature Medicine* 24 <https://doi.org/10.1038/s41591-018-0316-z> [accessed 4 September 2024.]

Andrew D Selbst and others, 'Fairness and Abstraction in Sociotechnical Systems' (2019) FAT\* '19 Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency 59 <https://doi.org/10.1145/3287560.3287598>

Angwin, J., Larson, J., Mattu, S. and Kirchner, L., (2016). Machine Bias. ProPublica. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 09 June 2024].

AU, 2020. Digital Transformation Strategy for Africa. Available at: <https://au.int/en/documents/20200518/digital-transformation-strategy-africa-2020-2030> [Accessed 11 June 2024].

Bagheri, A. and Abdollahi, M., (2021). AI in Africa: The Role of International Cooperation. Journal of AI Research, 64, pp.1-17.

Barocas, S. and Selbst, A.D., (2016). Big Data's Disparate Impact. California Law Review, 104(3), pp.671-732.

Barocas, S., Hardt, M. and Narayanan, A., (2019). Fairness and Machine Learning. MIT Press.

Bell M, *Anti-Discrimination Law and the European Union* (Oxford University Press 2002)

Berger, R., (2015). Now I see it, now I don’t: Researcher’s position and reflexivity in qualitative research. Qualitative Research, 15(2), pp.219-234.

Bietti E, 'From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy' (2020) 3(2) *Minds and Machines* 1

Binns, R., (2018). Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency, pp.149-159.

Bogen M and Rieke A, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias* (Upturn 2018) https://www.upturn.org/reports/2018/hiring-algorithms/ accessed 4 September 2024

Bowen, G.A., (2009). Document analysis as a qualitative research method. Qualitative Research Journal, 9(2), pp.27-40.

Brayne, S., (2017). Big Data Surveillance: The Case of Policing. American Sociological Review, 82(5), pp.977-1008.

Bryman, A., (2016). Social Research Methods. 5th ed. Oxford: Oxford University Press.

Buolamwini, J. and Gebru, T., (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. PMLR, pp.77-91.

Burrell, J., (2016). How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms. Big Data & Society, 3(1), pp.1-12.

CAC, 2021. Personal Information Protection Law. [pdf] Available at: <http://www.cac.gov.cn/2021-08/20/c_1631049986653198.html> [Accessed 22 May 2024].

Caliskan, A., Bryson, J.J. and Narayanan, A., (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), pp.183-186.

Calo, R., (2017). Artificial Intelligence Policy: A Primer and Roadmap. UC Davis Law Review, 51(2), pp.399-435.

Case C-645/19 Facebook Ireland Ltd v Data Protection Commissioner [2020] ECLI:EU:C: 2020:1042

Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Publishing Group 2016).

Cavoukian, A., (2010). Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario. Available at: <https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf> [Accessed 12 June 2024].

Chen, M., Mao, S. and Liu, Y., (2014). Big Data: A Survey. Mobile Networks and Applications, 19(2), pp.171-209.

Cheney-Lippold, J., (2017). We Are Data: Algorithms and The Making of Our Digital Selves. NYU Press.

Coeckelbergh M, *AI Ethics* (MIT Press 2020)

Council Directive 2000/43/EC of 29 June 2000 Implementing the Principle of Equal Treatment Between Persons Irrespective of Racial or Ethnic Origin [2000] OJ L180/22 (Racial Equality Directive).

Council Directive 2000/78/EC of 27 November 2000 Establishing a General Framework for Equal Treatment in Employment and Occupation [2000] OJ L303/16 (Employment Equality Directive).

Council of Europe, 'Recommendation CM/Rec (2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems' (8 April 2020)

Craig P and de Búrca G, *EU Law: Text, Cases, and Materials* (7th edn, Oxford University Press 2020)

Crawford K, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021)

Creswell, J.W. and Plano Clark, V.L., (2018). Designing and Conducting Mixed Methods Research. 3rd ed. Thousand Oaks, CA: Sage.

Dastin, J., (2018). Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women. Reuters. Available at: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> [Accessed 12 June 2024].

Data Protection Act 2018 (UK)

Dignum V, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way* (Springer 2019)

Domingos, P., (2015). The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books.

Doshi-Velez, F. and Kim, B., (2017). Towards a Rigorous Science of Interpretable Machine Learning. arXiv preprint arXiv:1702.08608.

Edwards L and Veale M, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2017) 16(3) *IEEE Security & Privacy* 46

Elgammal, A., (2019). AI Is Blurring the Definition of Artist. The Atlantic. Available at: <https://www.theatlantic.com/entertainment/archive/2019/01/ai-art-implications-creative-world/580054/> [Accessed 12 June 2024].

Engstrom, D.F. and Ho, D.E., (2020). AI in the Administrative State. The Regulatory Review. Available at: <https://www.theregreview.org/2020/01/30/engstrom-ho-ai-administrative-state/>

Engstrom, D.F. and Ho, D.E., (2020). Algorithmic Accountability in the Administrative State. Yale Journal on Regulation, 37(3), pp.800-861.

EPO, 2020. Decision of the Legal Board of Appeal of 27 January 2020 in Case J 8/20. European Patent Office. Available at: <https://www.epo.org/law-practice/case-law-appeals/recent/j200008eu1.html> [Accessed 07 June 2024].

Equality Act 2010 (UK).

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press, 2018. ISBN: 978-1-250-07431-7.

European Commission, 2019. GDPR: Google Fined €50 Million by the French Data Protection Authority on 21 January 2019. European Commission. Available at: <https://ec.europa.eu/newsroom/article29/items/612080/en> [Accessed 12 June 2024].

European Commission, 2020. EU Anti-Discrimination Directives. [pdf] Available at: <https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination_en> [Accessed 12 June 2024].

European Commission, 'Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM (2021) 206 final). [pdf] Available at: <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=75788>

European Parliament and Council, 2016. 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). Official Journal of the European Union, L119, pp.1-88.

European Union Agency for Fundamental Rights. *Getting the Future Right – Artificial Intelligence and Fundamental Rights*. Luxembourg: Publications Office of the European Union, 2020. Available at: <https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights>

Federal Trade Commission (FTC), 2020. Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues. [pdf] Available at: <https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report> [Accessed 12 June 2024].

Federal Trade Commission (FTC), 'Using Artificial Intelligence and Algorithms' (2021) <https://www.ftc.gov/> accessed 4 September 2024

Floridi, L. and Cowls, J., (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review. Available at: <https://doi.org/10.1162/99608f92.8cd550d1> [Accessed 17 June 2024].

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F. and Schafer, B., (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), pp.689-707.

Fowler, F.J., (2014). Survey Research Methods. 5th ed. Thousand Oaks, CA: Sage.

Fredman S, *Discrimination Law* (2nd edn, Oxford University Press 2011)

FTC, (2020). Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues. [pdf] Available at: <https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report> [Accessed 17 June 2024].

Gebru T and others, 'Datasheets for Datasets' (2021) 64(12) *Communications of the ACM* 86

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Daumé III, H. and Crawford, K., (2018). Datasheets for Datasets. arXiv preprint arXiv:1803.09010.

Ghassemi M and others, 'A Review of Challenges and Opportunities in Machine Learning for Health' (2020) 21 *AMIA Annual Symposium Proceedings Archive* 191

Ghassemi, M., Oakden-Rayner, L. and Beam, A.L., (2020). The False Hope of Current Approaches to Explainable Artificial Intelligence in Health Care. The Lancet Digital Health, 2(8), pp. e425-e430.

Gless, S., Silverman, E. and Weigend, T., (2016). If Robots Cause Harm, Who is to Blame? Self-driving Cars and Criminal Liability. New Criminal Law Review: An International and Interdisciplinary Journal, 19(3), pp.412-436.

Goodfellow, I., Bengio, Y. and Courville, A., (2016). Deep Learning. MIT Press.

Goodman, B. and Flaxman, S., (2017). European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”. AI Magazine, 38(3), pp.50-57.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. and Pedreschi, D., (2018). A Survey of Methods for Explaining Black Box Models. ACM Computing Surveys (CSUR), 51(5), pp.1-42.

Hardt, M., Price, E. and Srebro, N., (2016). Equality of Opportunity in Supervised Learning. In: Advances in Neural Information Processing Systems (pp. 3315-3323).

Harini Suresh and John V Guttag, 'A Framework for Understanding Unintended Consequences of Machine Learning' (2021) 64(9) *Communications of the ACM* 62 <https://doi.org/10.1145/3442188> [accessed 4 September 2024.]

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M. and Wallach, H., (2019). Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp.1-16.

House of Representatives, 2019. Algorithmic Accountability Act of 2019. [online] Available at: <https://www.congress.gov/bill/116th-congress/house-bill/2231> [Accessed 09 June 2024].

Hurley, M. and Adebayo, J., (2016). Credit Scoring in the Era of Big Data. Yale Journal of Law and Technology, 18(1), pp.148-216.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning* (MIT Press 2016).

IEEE, 2019. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. [pdf] Available at: <https://ethicsinaction.ieee.org> [Accessed 12 June 2024].

Information Commissioner's Office (ICO), 2020. Explaining decisions made with AI. [pdf] ICO. Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-ai/> [Accessed 12 June 2024].

Inioluwa Deborah Raji and Joy Buolamwini, 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products' (2019) AIES '19 Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society 429 <https://doi.org/10.1145/3306618.3314244>

International Conference of Data Protection and Privacy Commissioners (ICDPPC), 2020. International Cooperation. [online] Available at: <https://icdppc.org/> [Accessed 02 June 2024].

Isaak, J. and Hanna, M.J., (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. Computer, 51(8), pp.56-59.

ISO, 2020. ISO/IEC JTC 1/SC 42 Artificial Intelligence. Available at: <https://www.iso.org/committee/6794475.html> [Accessed 12 June 2024].

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H. and Wang, Y., (2017). Artificial Intelligence in Healthcare: Past, Present and Future. Stroke and Vascular Neurology, 2(4), pp.230-243.

Justice Committee, 2020. Post Office Horizon IT Inquiry. UK Parliament. Available at: <https://committees.parliament.uk/work/306/post-office-horizon-it-inquiry/> [Accessed 12 June 2024].

Kearns M and Roth A, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* (Oxford University Press 2019)

Kehl, D., Guo, P. and Kessler, S., (2017). Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Harvard Law School. Available at: <https://dash.harvard.edu/handle/1/33746041> [Accessed 12 June 2024].

Kosinski, M., Stillwell, D. and Graepel, T., 2013. Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 110(15), pp.5802-5805.

Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G. and Yu, H., (2017). Accountable algorithms. University of Pennsylvania Law Review, 165(3), pp.633-705.

Kuner, C., (2020). The GDPR and Transatlantic Data Flows. International Data Privacy Law, 10(2), pp.112-122.

LeCun, Y., Bengio, Y. and Hinton, G., (2015). Deep Learning. Nature, 521(7553), pp.436-444.

Lum, K. and Isaac, W., (2016). To predict and serve? In: Significance, 13(5), pp.14-19.

Lundberg, S.M. and Lee, S.I., (2017). A Unified Approach to Interpreting Model Predictions. In: Advances in Neural Information Processing Systems (pp. 4765-4774).

Lynskey O, *The Foundations of EU Data Protection Law* (Oxford University Press 2015)

MacFeely, S., (2019). The Big (Data) Bang: Opportunities and Challenges for Compiling SDG Indicators. Global Policy, 10(S1), pp.121-133.

Mantelero, A., (2018). AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. Computer Law & Security Review, 34(4), pp.754-772.

McCrudden C, *Equality and Discrimination Law in the UK* (Oxford University Press 2016)

Michael Kearns and Aaron Roth, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* (Oxford University Press 2019).

Miranda Bogen and Aaron Rieke, 'Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias' (Upturn, 2018) <https://www.upturn.org/reports/2018/hiring-algorithms/> [accessed 4 September 2024.]

Mitchell, M., Potash, E., Barocas, S., D'Amour, A. and Lum, K., (2019). Algorithmic Fairness: Choices, Assumptions, and Definitions. Annual Review of Statistics and Its Application, 6(1), pp.611-635.

Mittelstadt B, 'Principles Alone Cannot Guarantee Ethical AI' (2019) 1 *Nature Machine Intelligence* 501

Mittelstadt B, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 *Big Data & Society* 1

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L., (2016). The Ethics of Algorithms: Mapping the Debate. Big Data & Society, 3(2), pp.1-21.

Moussly, R., (2018). Marriott's data breach and the importance of robust cybersecurity. Forbes. Available at: <https://www.forbes.com/sites/forbestechcouncil/2018/12/14/marriotts-data-breach-and-the-importance-of-robust-cybersecurity/> [Accessed 12 June 2024].

Ni Loideain N and Adams R, 'From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of EU Data Protection Law' (2020) 11(2) *European Journal of Law and Technology* 3

Nicholas Diakopoulos, *Automating the News: How Algorithms Are Rewriting the Media* (Harvard University Press 2019).

Ninareh Mehrabi and others, 'A Survey on Bias and Fairness in Machine Learning' (2021) 54(6) *ACM Computing Surveys* 115 <https://doi.org/10.1145/3457607>

NIST, 2021. AI Risk Management Framework. Available at: <https://www.nist.gov/itl/ai-risk-management-framework> [Accessed 12 June 2024].

Noble, S.U., (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

Numhauser-Henning A and Laulom S, 'Harassment Related to Sex and Sexual Harassment Law in 33 European Countries' (2012) European Network of Legal Experts in the Field of Gender Equality Report

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting racial bias in an algorithm used to manage the health of populations." *Science* 366, no. 6464 (2019): pp.447-453.

O'Cinneide C, 'The Race Directive: A Shadow Report' (2005) 5(1) *European Anti-Discrimination Law Review* 15

Ohm, P., (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA Law Review, 57(6), pp.1701-1777.

Orb, A., Eisenhauer, L. and Wynaden, D., (2001). Ethics in qualitative research. Journal of Nursing Scholarship, 33(1), pp.93-96.

Organisation for Economic Co-operation and Development (OECD), 'Recommendation of the Council on Artificial Intelligence' OECD/LEGAL/0449 (22 May 2019

Pagallo, U., (2013). The Laws of Robots: Crimes, Contracts, and Torts. Springer.

Pasquale F, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015)

Powles, J. and Hodson, H., (2017). Google DeepMind and Healthcare in an Age of Algorithms. Health and Technology, 7(4), pp.351-367.

*R (on the application of Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058, [2020] 1 WLR 672

R (on the application of Edward Santon) v Secretary of State for Work and Pensions [2020] EWHC 381 (Admin), [2020] PTSR 1562

Raji, I.D. and Buolamwini, J., (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. ACM, pp.429-435.

Rashida Richardson, Jason Schultz, and Kate Crawford, 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice' (2019) 94(2) *NYU Law Review* 192 <https://www.nyulawreview.org/issues/volume-94-number-2/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/> [accessed 4 September 2024]

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) [2016] OJ L119/1 (GDPR).

Reisman, D., Schultz, J., Crawford, K. and Whittaker, M., (2018). Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability. AI Now Institute. Available at: <https://ainowinstitute.org/aiareport2018.pdf> [Accessed 12 June 2024].

Ribeiro, M.T., Singh, S. and Guestrin, C., (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).

Richardson, R., Schultz, J.M. and Crawford, K., (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. New York University Law Review Online, 94, pp.192-233.

Russell, S.J. and Norvig, P., (2016). Artificial Intelligence: A Modern Approach. 3rd ed. Pearson.

Russell, S.J. and Norvig, P., (2020). Artificial Intelligence: A Modern Approach. 4th ed. Pearson.

Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (NYU Press 2018).

Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) *International Data Privacy Law* 76 <https://doi.org/10.1093/idpl/ipx005>

Selbst AD and Powles J, 'Meaningful Information and the Right to Explanation' (2017) 7(4) *International Data Privacy Law* 233

Selbst AD, 'Disparate Impact in Big Data Policing' (2017) 52(1) *Georgia Law Review* 109

Shao, K., (2020). Artificial Intelligence and Intellectual Property. China IP Magazine. Available at: <http://www.chinaipmagazine.com/en/journal-show.asp?id=1620> [Accessed 12 June 2024].

Shin, D., (2020). The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI. International Journal of Human-Computer Studies, 140, p.102456.

Smith, H., (2020). The Challenges of Implementing AI Ethical Guidelines. Journal of AI and Ethics, 1(1), pp.43-54.

Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104(3) *California Law Review* 671 <https://doi.org/10.15779/Z38BG31>

Solove, D.J. and Schwartz, P.M., (2020). Information Privacy Law. 7th ed. Wolters Kluwer.

Srinivasan, R., (2018). Lessons from the Equifax hack. MIT Technology Review. Available at: <https://www.technologyreview.com/2018/09/10/240875/lessons-from-the-equifax-hack/> [Accessed 12 June 2024].

Stark, L. and Hoffmann, A.L., (2019). Data Is the New What? Popular Metaphors & Professional Ethics in Emerging Data Cultures. Journal of Cultural Analytics, 4(2).

Stuart J Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (4th edn, Pearson 2021).

Tene, O. and Polonetsky, J., (2013). Big Data for All: Privacy and User Control in the Age of Analytics. Northwestern Journal of Technology and Intellectual Property, 11(5), pp.239-273.

Thrun, S., (2010). Towards Robotic Cars. Communications of the ACM, 53(4), pp.99-106.

Timnit Gebru and others, 'Datasheets for Datasets' (2021) 64(12) *Communications of the ACM* 86 <https://doi.org/10.1145/3458723>

Topol, E.J., (2019). High-performance Medicine: The Convergence of Human and Artificial Intelligence. Nature Medicine, 25(1), pp.44-56.

UK Government, (2010). Equality Act 2010. [online] Available at: <https://www.legislation.gov.uk/ukpga/2010/15/contents> [Accessed 12 June 2024].

UK Government, (2018). Data Protection Act 2018. [online] Available at: <https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted> [Accessed 12 June 2024].

UK Government, (2020). National AI Strategy. Available at: <https://www.gov.uk/government/publications/national-ai-strategy> [Accessed 12 June 2024].

UK Government, 'Data Protection and Digital Information Bill' (2022) https://services.parliament.uk/Bills/2022/dataprotectionanddigitalinformation.html accessed 4 September 2024.

UK Government, 'National AI Strategy' (2021) https://www.gov.uk/government/publications/national-ai-strategy accessed 4 September 2024.

UNESCO, (2021). Recommendation on the Ethics of Artificial Intelligence. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000373434> [Accessed 12 June 2024].

Van den Hoven, J., (2020). Ethics and the Governance of Artificial Intelligence. Humanities and Social Sciences Communications, 7(1), pp.1-6.

Vayena, E., Blasimme, A. and Cohen, I.G., (2018). Machine Learning in Medicine: Addressing Ethical Challenges. PLoS Medicine, 15(11), p.e1002689.

Veale, M. and Borgesius, F.Z., (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), pp.97-112.

Veale, M. and Edwards, L., (2018). Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling. Computer Law & Security Review, 34(2), pp.398-404.

Veale, M., Van Kleek, M. and Binns, R., (2018). Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, pp.1-14.

Wachter, S., Mittelstadt, B. and Floridi, L., (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), pp.76-99.

Wachter, Sandra, Brent Mittelstadt, and Chris Russell. "Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI." *Computer Law & Security Review* 41 (2021): 105567. DOI: [10.1016/j.clsr.2021.105567](http://www.10.1016/j.clsr.2021.105567.com)

Weber, R.H., (2010). Internet of Things–New Security and Privacy Challenges. Computer Law & Security Review, 26(1), pp.23-30.

Westen P, *Speaking of Equality: An Analysis of the Rhetorical Force of "Equality" in Moral and Legal Discourse* (Princeton University Press 2014)

Wiles, R., Crow, G., Heath, S. and Charles, V., (2008). The management of confidentiality and anonymity in social research. International Journal of Social Research Methodology, 11(5), pp.417-428.

Yin, R.K., (2018). Case Study Research and Applications: Design and Methods. 6th ed. Thousand Oaks, CA: Sage.

Zarsky, T.Z., (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. Science, Technology, & Human Values, 41(1), pp.118-132.

Zarsky, T.Z., (2017). Incompatible: The GDPR in the Age of Big Data. Seton Hall Law Review, 47(4), pp.995-1020.

Zemel, R.S., Wu, Y., Swersky, K., Pitassi, T. and Dwork, C., (2013). Learning Fair Representations. In: Proceedings of the 30th International Conference on Machine Learning (pp. 325-333).

Zliobaite, I. (2017). Measuring Discrimination in Algorithmic Decision Making. Data Mining and Knowledge Discovery, 31(4), 1060-1089.

Zuboff S, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs 2019)

**List of Abbreviations**

| **Abbreviation** | **Full Term** |
| --- | --- |
| AI | Artificial Intelligence |
| ADM | Automated Decision-Making |
| GDPR | General Data Protection Regulation |
| DWP | Department for Work and Pensions |
| HRA | Human Rights Act 1998 |
| ECHR | European Convention on Human Rights |
| ADMS | Automated Decision-Making Systems |
| DPA | Data Protection Act |
| EU | European Union |
| UK | United Kingdom |
| OECD | Organisation for Economic Co-operation and Development |
| UN | United Nations |

1. European Union Agency for Fundamental Rights. "Getting the Future Right – Artificial Intelligence and Fundamental Rights." (2020). [↑](#footnote-ref-1)
2. Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting racial bias in an algorithm used to manage the health of populations." *Science* 366, no. 6464 (2019): 447-453. [↑](#footnote-ref-2)
3. Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* St. Martin's Press, 2018. [↑](#footnote-ref-3)
4. Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed. Pearson, 2021. ISBN: 9780134610993. [↑](#footnote-ref-4)
5. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. ISBN: 9780262035613. [↑](#footnote-ref-5)
6. Suresh, Harini, and John V. Guttag. "A Framework for Understanding Unintended Consequences of Machine Learning." *Communications of the ACM* 64, no. 9 (2021): 62-71. DOI: 10.1145/3442188 [↑](#footnote-ref-6)
7. Esteva, Andre, et al. "A Guide to Deep Learning in Healthcare." *Nature Medicine* 25 (2019): 24-29. DOI: 10.1038/s41591-018-0316-z [↑](#footnote-ref-7)
8. Richardson, Rashida, Jason Schultz, and Kate Crawford. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *NYU Law Review* 94, no. 2 (2019): 192-233. Available at: https://www.nyulawreview.org/issues/volume-94-number-2/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/ [↑](#footnote-ref-8)
9. Bogen, Miranda, and Aaron Rieke. "Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias." Upturn, 2018. Available at: https://www.upturn.org/reports/2018/hiring-algorithms/ [↑](#footnote-ref-9)
10. Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." *California Law Review* 104, no. 3 (2016): 671-732. DOI: 10.15779/Z38BG31 [↑](#footnote-ref-10)
11. Selbst, Andrew D., et al. "Fairness and Abstraction in Sociotechnical Systems." *Proceedings of the Conference on Fairness, Accountability, and Transparency* (FAT\* '19), January 2019. DOI: 10.1145/3287560.3287598 [↑](#footnote-ref-11)
12. Miranda Bogen and Aaron Rieke, 'Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias' *Upturn* (2018) <https://www.upturn.org/reports/2018/hiring-algorithms/> [↑](#footnote-ref-12)
13. Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104(3) *California Law Review* 671. [↑](#footnote-ref-13)
14. Marzyeh Ghassemi and others, 'A Review of Challenges and Opportunities in Machine Learning for Health' (2020) 21 *AMIA Annual Symposium Proceedings Archive* 191 [↑](#footnote-ref-14)
15. Brent Mittelstadt, 'Principles Alone Cannot Guarantee Ethical AI' (2019) 1 *Nature Machine Intelligence* 501. [↑](#footnote-ref-15)
16. 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) [2016] OJ L119/1. [↑](#footnote-ref-16)
17. Equality Act 2010 (UK) [↑](#footnote-ref-17)
18. *R (on the application of Edward Santon) v Secretary of State for Work and Pensions* [2020] EWHC 381 (Admin), [2020] PTSR 1562 [↑](#footnote-ref-18)
19. *R (on the application of Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058, [2020] 1 WLR 672 [↑](#footnote-ref-19)
20. Sarah Brayne, 'Big Data Surveillance: The Case of Policing' (2017) 82(5) *American Sociological Review* 977 [↑](#footnote-ref-20)
21. Virginia Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way* (Springer 2019). [↑](#footnote-ref-21)
22. Brent Mittelstadt, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 *Big Data & Society* 1 [↑](#footnote-ref-22)
23. Richard Zemel and others, 'Learning Fair Representations' (2013) 28 *Proceedings of the International Conference on Machine Learning* 325. [↑](#footnote-ref-23)
24. Inioluwa Deborah Raji and Joy Buolamwini, 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products' (2019) *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* 429. [↑](#footnote-ref-24)
25. Timnit Gebru and others, 'Datasheets for Datasets' (2021) 64(12) *Communications of the ACM* 86. [↑](#footnote-ref-25)
26. Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2017) 16(3) *IEEE Security & Privacy* 46. [↑](#footnote-ref-26)
27. Mark Bell, *Anti-Discrimination Law and the European Union* (Oxford University Press 2002) [↑](#footnote-ref-27)
28. Paul Craig and Gráinne de Búrca, *EU Law: Text, Cases, and Materials* (7th edn, Oxford University Press 2020) 354-355. [↑](#footnote-ref-28)
29. Christopher McCrudden, *Equality and Discrimination Law in the UK* (Oxford University Press 2016). [↑](#footnote-ref-29)
30. Racial Equality Directive (2000/43/EC) art 2 [↑](#footnote-ref-30)
31. Ibid art 8 [↑](#footnote-ref-31)
32. Employment Equality Directive (2000/78/EC) art 5 [↑](#footnote-ref-32)
33. General Data Protection Regulation (EU) 2016/679, art 22 [↑](#footnote-ref-33)
34. Ibid art 9 [↑](#footnote-ref-34)
35. European Commission, 'Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' COM (2021) 206 final. [↑](#footnote-ref-35)
36. Ibid [↑](#footnote-ref-36)
37. *Equality Act 2010* (UK) [↑](#footnote-ref-37)
38. Sandra Fredman, *Discrimination Law* (2nd edn, OUP 2011) [↑](#footnote-ref-38)
39. *Data Protection Act 2018* (UK) [↑](#footnote-ref-39)
40. *Data Protection Act 2018* (UK), s 10 [↑](#footnote-ref-40)
41. UK Government, 'National AI Strategy' (2021) <https://www.gov.uk/government/publications/national-ai-strategy> [↑](#footnote-ref-41)
42. UK Government, 'Data Protection and Digital Information Bill' (2022) <https://services.parliament.uk/Bills/2022/dataprotectionanddigitalinformation.html> [↑](#footnote-ref-42)
43. *R (on the application of Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058, [2020] 1 WLR 672 [↑](#footnote-ref-43)
44. Rechtbank Den Haag, 5 februari 2020, ECLI:NL: RBDHA:2020:865 (SyRI) [↑](#footnote-ref-44)
45. *R (on the application of Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058 [↑](#footnote-ref-45)
46. Ibid [↑](#footnote-ref-46)
47. *R (on the application of Edward Santon) v Secretary of State for Work and Pensions* [2020] EWHC 381 (Admin), [2020] PTSR 1562 [↑](#footnote-ref-47)
48. Council Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L180/22 (Racial Equality Directive). [↑](#footnote-ref-48)
49. Equality Act 2010 (UK) [↑](#footnote-ref-49)
50. 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) [2016] OJ L119/1, art 22 [↑](#footnote-ref-50)
51. Ibid, art 9 [↑](#footnote-ref-51)
52. *R (on the application of Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058, [2020] 1 WLR 672 [↑](#footnote-ref-52)
53. Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2017) 16(3) *IEEE Security & Privacy* 46. [↑](#footnote-ref-53)
54. Andrew D Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7(4) *International Data Privacy Law* 233. [↑](#footnote-ref-54)
55. Wachter S, Mittelstadt B and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) *International Data Privacy Law* 76. [↑](#footnote-ref-55)
56. Sandra Wachter, Brent Mittelstadt, and Chris Russell, 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI' (2021) 41 *Computer Law & Security Review* 105567. [↑](#footnote-ref-56)
57. European Commission, 'Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM (2021) 206 final. [↑](#footnote-ref-57)
58. UK Government, 'National AI Strategy' (2021) <https://www.gov.uk/government/publications/national-ai-strategy> [↑](#footnote-ref-58)
59. Nóra Ni Loideain and Rachel Adams, 'From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of EU Data Protection Law' (2020) 11(2) *European Journal of Law and Technology* 3. [↑](#footnote-ref-59)
60. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015). [↑](#footnote-ref-60)
61. Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2017) 16(3) *IEEE Security & Privacy* 46. [↑](#footnote-ref-61)
62. *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) [2016] OJ L119/1* art 22. [↑](#footnote-ref-62)
63. Inioluwa Deborah Raji and Joy Buolamwini, 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products' (2019) AIES '19 Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society 429. [↑](#footnote-ref-63)
64. Inioluwa Deborah Raji and Joy Buolamwini, 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products' (2019) AIES '19 Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society 429. [↑](#footnote-ref-64)
65. European Commission, 'Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM (2021) 206 final. [↑](#footnote-ref-65)
66. European Commission, 'Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM (2021) 206 final. [↑](#footnote-ref-66)
67. Wachter S, Mittelstadt B and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) *International Data Privacy Law* 76. [↑](#footnote-ref-67)
68. Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104(3) *California Law Review* 671. [↑](#footnote-ref-68)
69. Timnit Gebru and others, 'Datasheets for Dataset’s (2021) 64(12) *Communications of the ACM* 86 [↑](#footnote-ref-69)
70. *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) [2016] OJ L119/1* art 9. [↑](#footnote-ref-70)
71. Wachter S, Mittelstadt B and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) *International Data Privacy Law* 76. [↑](#footnote-ref-71)
72. Sandra Wachter, Brent Mittelstadt, and Chris Russell, 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI' (2021) 41 *Computer Law & Security Review* 105567. [↑](#footnote-ref-72)
73. Michael Kearns and Aaron Roth, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* (Oxford University Press 2019). [↑](#footnote-ref-73)
74. Nóra Ni Loideain and Rachel Adams, 'From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of EU Data Protection Law' (2020) 11(2) *European Journal of Law and Technology* 3. [↑](#footnote-ref-74)
75. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015). [↑](#footnote-ref-75)
76. Andrew D Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7(4) *International Data Privacy Law* 233. [↑](#footnote-ref-76)
77. Organisation for Economic Co-operation and Development (OECD), 'Recommendation of the Council on Artificial Intelligence' OECD/LEGAL/0449 (22 May 2019). [↑](#footnote-ref-77)
78. Council of Europe, 'Recommendation CM/Rec (2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems' (8 April 2020). [↑](#footnote-ref-78)