



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Algorithmic Discrimination in Europe

Citation for published version:

Gerards, J & Xenidis, R 2021, *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law*. European Commission. <<https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1>>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





European
Commission

European network of legal experts in
gender equality and non-discrimination



Algorithmic discrimination in Europe

Challenges and opportunities
for gender equality and
non-discrimination law

Including summaries in
English, French and German

*Justice
and Consumers*

EUROPEAN COMMISSION

Directorate-General for Justice and Consumers
Directorate D — Equality and Union citizenship
Unit D.1 Non-discrimination and Roma coordination
Unit D.2 Gender Equality

European Commission
B-1049 Brussels

Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law

A special report

Authors

Janneke Gerards (Utrecht University) and Raphaële Xenidis
(University of Edinburgh and University of Copenhagen)

2020

The authors would like to warmly thank the whole coordination team of the European Network of Legal Experts in gender equality and non-discrimination at Utrecht University and in particular Birte Böök, Franka van Hoof, Alexandra Timmer and Linda Senden, the executive team at Human European Consultancy, in particular Yvonne van Leeuwen-Lohde and Marcel Zwamborn, the 31 national experts in gender equality law and the national experts in non-discrimination law who collaborated on this report for their invaluable help, as well as Florianne Peters van Neijenhof and Pita Klaassen for their excellent research support. The drafting of this report was concluded in October 2020.

***Europe Direct is a service to help you find answers
to your questions about the European Union.***

Freephone number (*):

00 800 6 7 8 9 10 11

(*)The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

LEGAL NOTICE

This document has been prepared for the European Commission; however, it reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

More information on the European Union is available on the Internet (<http://www.europa.eu>).

Luxembourg: Publications Office of the European Union, 2021

ISBN 978-92-76-20746-7

doi:10.2838/544956

Catalogue number DS-02-20-549-EN-N

© European Union, 2021

Contents

EXECUTIVE SUMMARY	7
RÉSUMÉ	13
ZUSAMMENFASSUNG	20
GENERAL INTRODUCTION	27
Subject, context and scope of the report	27
Methodology	29
Structure	29
1 WHAT IS ALGORITHMIC DISCRIMINATION AND WHAT IS NEW ABOUT IT?	31
1.1 Introduction	31
1.2 Types of algorithms	32
1.2.1 Rule-based algorithms	32
1.2.2 Machine-learning algorithms	33
1.2.3 Deep learning	36
1.2.4 Enabling technologies and combining algorithms: AI	36
1.3 Stages of algorithmic decision making and uses of algorithms	37
1.3.1 Planning stage	38
1.3.2 Development stage	39
1.3.3 Decision-making and use stage	39
1.4 Algorithmic characteristics and challenges	40
1.4.1 The human factor and the stereotyping and cognitive bias challenge	41
1.4.2 The data challenge	42
1.4.3 The correlation and proxies challenge	44
1.4.4 The transparency and explainability challenge	45
1.4.5 The scale and speed challenge	46
1.4.6 The responsibility challenge	46
1.5 Terminology and interactions between gender equality and non-discrimination law and data protection law	47
1.5.1 Terminology: ‘bias’ and ‘fairness’ versus ‘discrimination’ and ‘equality’	47
1.5.2 Interactions between non-discrimination law and data protection	48
1.6 Conclusion	50
2 CHALLENGES TO THE EU GENDER EQUALITY AND NON-DISCRIMINATION LEGAL FRAMEWORK	53
2.1 The scope of EU gender equality and non-discrimination law in light of the problem of algorithmic discrimination	53
2.1.1 The legal framework	54
2.1.2 Equal pay, employment and self-employment	55
2.1.3 Goods and services: problematic gaps in the material scope	58
2.2 Protected grounds and algorithmic discrimination	62
2.2.1 Algorithmic gender-based classification	62
2.2.2 Correlations and proxies	63
2.2.3 New forms and grounds of discrimination	64
2.2.4 The CJEU’s interpretation of the grounds listed in Article 21 of the Charter of Fundamental Rights	65
2.2.5 Algorithmic granularity and intersectionality	65
2.2.6 The dynamic nature of algorithmic categorisations	66
2.3 The types of discrimination defined in EU law	67
2.3.1 Direct discrimination: an uneasy fit with algorithmic discrimination	67

2.3.2	Indirect discrimination: a better conceptual fit with a wide pool of potential justifications	70
2.4	Questions of proof, responsibility and liability	73
2.5	Conclusion	75
3	CHALLENGES FOR THE EUROPEAN STATES IN RELATION TO ALGORITHMIC DISCRIMINATION	78
3.1	Examples of the use of algorithms in European countries	78
3.1.1	Introduction	78
3.1.2	Examples of the use of algorithms in the public sector	78
3.1.3	Examples of use of algorithms in the private sector	83
3.2	Problems related to algorithmic decision-making	85
3.2.1	Biases in data	85
3.2.2	Discriminatory effects	86
3.2.3	Transparency problems and lack of information	87
3.2.4	Detecting algorithmic discrimination	87
3.2.5	Responsibility issues	88
3.2.6	A gender digital gap in European countries	88
3.3	Awareness of risks of algorithmic discrimination in European countries	90
3.3.1	Public discussions on the impact of algorithms on gender equality and non-discrimination	90
3.3.2	Scientific discussions on the impact of algorithms on gender equality and non-discrimination	98
3.4	Legal responses to algorithmic discrimination in the European countries	107
3.4.1	Legislative instruments	107
3.4.2	(Semi-)judicial application and enforcement of legislation	113
3.5	Conclusion	117
4	ENFORCING ALGORITHMIC EQUALITY: SOLUTIONS AND OPPORTUNITIES FOR GENDER EQUALITY AND NON-DISCRIMINATION	121
4.1	Introduction	121
4.2	Benefits and opportunities of algorithmic decision making	121
4.3	Tackling algorithmic discrimination: a review of national good practice in European countries	124
4.3.1	Monitoring algorithmic discrimination: examples of good practices and opportunities	124
4.3.2	Addressing algorithmic discrimination: examples of good practices and opportunities	125
4.3.3	The diversity question in relevant professional and educational communities	133
4.4	Potential solutions and tools to prevent and remedy algorithmic discrimination: a tridimensional approach	140
4.4.1	Introduction	140
4.4.2	Legal solutions	141
4.4.3	Knowledge-based solutions	146
4.4.4	Technology-based solutions	148
4.5	Conclusion: PROTECT – proposal for an integrated approach to algorithmic discrimination	150
	GENERAL CONCLUSIONS	152
	BIBLIOGRAPHY	153
	ANNEX – QUESTIONNAIRE ALGORITHMIC DISCRIMINATION IN EUROPE: CHALLENGES AND OPPORTUNITIES FOR GENDER EQUALITY AND NON-DISCRIMINATION LAW	183

Members of the European network of legal experts in gender equality and non-discrimination

Management team

General coordinator	Marcel Zwamborn	Human European Consultancy
Specialist coordinator gender equality law	Linda Senden	Utrecht University
Content coordinator gender equality law	Alexandra Timmer	Utrecht University
Specialist coordinator non-discrimination law	Isabelle Chopin	Migration Policy Group
Project managers	Ivette Groenendijk Yvonne van Leeuwen-Lohde	Human European Consultancy
Content managers gender equality law	Franka van Hoof Birte Böök	Utrecht University
Content manager non-discrimination law	Catharina Germaine	Migration Policy Group

Senior experts

Senior expert on gender equality law	Susanne Burri
Senior expert on age	Elaine Dewhurst
Senior expert on sexual orientation/trans/intersex people	Peter Dunne
Senior expert on racial or ethnic origin	Lilla Farkas
Senior expert on EU and human rights law	Christopher McCrudden
Senior expert on social security	Frans Pennings
Senior expert on religion or belief	Isabelle Rorive
Senior expert on EU law, CJEU case law, sex, gender identity and gender expression in relation to trans and intersex people	Christa Tobler
Senior expert on disability	Lisa Waddington

National experts

	Non-discrimination	Gender
Albania	Irma Baraku	Entela Baci
Austria	Dieter Schindlauer	Martina Thomasberger
Belgium	Emmanuelle Bribosia	Nathalie Wuiame
Bulgaria	Margarita Ilieva	Genoveva Tisheva
Croatia	Ines Bojić	Adrijana Martinović
Cyprus	Corina Demetriou	Vera Pavlou
Czechia	Jakub Tomšej	Kristina Koldinská
Denmark	Pia Justesen	Natalie Videbaek Munkholm
Estonia	Vadim Poleshchuk	Anu Laas
Finland	Rainer Hiltunen	Kevät Nousiainen
France	Sophie Latraverse	Marie Mercat-Bruns
Germany	Matthias Mahlmann	Ulrike Lembke
Greece	Athanasios Theodoridis	Panagiota Petroglou
Hungary	András Kádár	Lídia Hermína Balogh
Iceland	Gudrun D. Gudmundsdottir	Herdís Thorgeirsdóttir
Ireland	Judy Walsh	Frances Meenan
Italy	Chiara Favilli	Simonetta Renga
Latvia	Anhelita Kamenska	Kristīne Dupate
Liechtenstein	Patricia Hornich	Nicole Mathé
Lithuania	Birutė Sabatauskaitė	Tomas Davulis
Luxembourg	Tania Hoffmann	Nicole Kerschen
Malta	Tonio Ellul	Romina Bartolo
Montenegro	Maja Kostić-Mandić	Vesna Simovic-Zvicer
Netherlands	Karin de Vries	Marlies Vegter
North Macedonia	Biljana Kotevska	Biljana Kotevska
Norway	Lene Løvdal	Marte Bauge
Poland	Łukasz Bojarski	Eleonora Zielinska and Anna Cybulko
Portugal	Dulce Lopes and Joana Vicente	Maria do Rosário Palma Ramalho
Romania	Romanița Iordache	Iustina Ionescu
Serbia	Ivana Krstić Davinic	Ivana Krstić Davinic
Slovakia	Vanda Durbáková	Zuzana Magurová
Slovenia	Neža Kogovšek Šalamon	Tanja Koderman Sever
Spain	Lorenzo Cachón	Dolores Morondo Taramundi
Sweden	Paul Lappalainen	Jenny Julén Votinius
Turkey	Ulaş Karan	Kadriye Bakirci
United Kingdom	Lucy Vickers	Rachel Horton

Executive summary

In recent years, media stories and scholarly articles about algorithmic discrimination have flourished. In 2019, for example, the AppleCard algorithm was found to grant higher credit limits to men than to women despite the latter having higher credit scores.¹ In 2014, Amazon developed an algorithmic hiring prototype, which was later found to discriminate against women and had to be abandoned.² A recent empirical study showed how the targeting of online ads can reinforce stereotyping and segregation on the labour market: during the experiment, researchers used the Facebook advertising platform to neutrally disseminate various employment ads. In the end, cashier positions in supermarkets reached an audience composed of 85 % women, while ads for taxi driver positions reached a 75 % black audience and ads for lumberjack positions reached an audience that was 90 % male and 72 % white.³ A global consensus has emerged among both researchers and policy-makers that risks of algorithmic discrimination are pervasive and multifaceted. In this context, understanding these risks and the types of legal challenges they create is key to ensuring equality and combating discrimination.

Indeed, in 2019, the European Commission published a white paper on artificial intelligence, which recognised that the increasing use of algorithms in Europe poses specific risks in terms of fundamental rights protection and in particular in terms of equality and non-discrimination.⁴ Such risks are also recognised by the Commission's recent Gender Equality Strategy 2020-2025, which acknowledges that 'AI [...] risks intensifying gender inequalities'.⁵ In response, the EU has called for the creation of an 'ecosystem of trust', which demands that 'European AI is grounded in [EU] values and fundamental rights' among which the right to equality and non-discrimination is central.⁶

This report investigates how algorithmic discrimination challenges the set of legal guarantees put in place in Europe to combat discrimination and ensure equal treatment. More specifically, it examines whether and how the current gender equality and non-discrimination legislative framework in place in the EU can adequately capture and redress algorithmic discrimination. It explores the gaps and weaknesses that emerge at both the EU and national levels from the interaction between, on the one hand, the specific types of discrimination that arise when algorithms are used in decision-making systems and, on the other, the particular material and personal scope of the existing legislative framework. This report also maps out the existing legal solutions, accompanying policy measures and good practice to address and redress algorithmic discrimination both at EU and national levels. Moreover, this report proposes its own integrated set of legal, knowledge-based and technological solutions to the problem of algorithmic discrimination.

-
- 1 Gupta, AH (2019), 'Are Algorithms Sexist?', *The New York Times* (15 November 2019), available at: www.nytimes.com/2019/11/15/us/apple-card-goldman-sachs.html.
 - 2 Dastin, J (2018), 'Amazon scraps secret AI recruiting tool that showed bias against women' (10 October 2018), *Reuters*, available at: www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.
 - 3 The authors specify that these numbers correspond to 'the most extreme cases' of skewed distribution. In the experiment conducted, they selected an identical audience for all three adverts. Ali, M and others (2019), 'Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes', arXiv preprint, available at: arXiv:190402095 1.
 - 4 European Commission (2020), 'White Paper on Artificial Intelligence: A European approach to excellence and trust' COM(2020) 65 final (Brussels 2020), p. 3, p. 11.
 - 5 European Commission (2020) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'A Union of Equality: Gender Equality Strategy 2020-2025' COM(2020) 152 final (Brussels 2020), 6.
 - 6 European Commission (2020), 'White Paper on Artificial Intelligence', p. 2.

Algorithms and discrimination: what are we talking about?

The first chapter of the report sets the scene for the discussion of the various problems raised by algorithms in relation to gender equality and non-discrimination law. It offers key definitions as well as an introduction to the various uses that can be made of algorithms and the different types of algorithms currently in use. Since different algorithmic technologies pose different types of challenges for gender equality and non-discrimination law, it is important to differentiate between the various types and the specific issues they pose. Section 1.2 on types of algorithms is mainly addressed to readers who are not familiar with the different types of algorithmic technologies and their characteristics, hence readers who are familiar with these might skip section 1.2 and jump directly to section 1.3. This next section takes the reader through the various phases in which discrimination can creep into algorithms. From design to use, and from planning to development and decision-making, bias can impact algorithms in several ways.

Chapter 1 also explains how human prejudices and stereotypes as well as societal structural inequalities reflected in the data used to train algorithms can lead to discriminatory algorithms. Piecing together these different insights, the core section of Chapter 1 highlights six major challenges that algorithms pose to gender equality and non-discrimination law:

- 1) *the human factor and the stereotyping and cognitive bias challenge* describe how implicit biases, harmful stereotypes and discriminatory prejudices held by humans risk infecting the algorithms humans create and how automation and anchoring biases reinforce these risks;
- 2) *the data challenge* describes how data embodies the historically consolidated patterns of discrimination that structure society and how training algorithms with such biased data, or with incorrect, unrepresentative or unbalanced data, leads to the reproduction of structural inequalities by these algorithms;
- 3) *the correlation and proxies challenge* – the correlation challenge explains how algorithms might reify and further enact discriminatory correlations (e.g. gender might negatively correlate with work performance, not because of a causal relationship, but because women historically have been consistently evaluated more negatively than men for the same work performance)⁷ by treating them as causalities and using them as foundations for further decisions, recommendations or predictions, while the proxies challenge outlines how removing protected characteristics from the pool of available input variables is insufficient in light of learning algorithms' ability to detect proxies for these protected characteristics;
- 4) *the transparency and explainability challenge* refers to difficulties in monitoring and proving algorithmic discrimination in light of the opacity of certain types of algorithms (even for computer scientists) and the lack of information about their inner workings (especially when codes and data are proprietary);
- 5) *the scale and speed challenge* describes how algorithmic discrimination can 'spread' at a wider scale and a much quicker pace than 'human' discrimination since algorithms both speed up and scale up decision making; and
- 6) *the responsibility, liability and accountability challenge* is the difficulty of identifying who to hold responsible, liable and/or accountable for a discriminatory outcome in the context of complex human-machine relationships, given that so many different parties are involved in the design, commercialisation and use of algorithms.

Chapter 1 closes with a short discussion of the terminology adopted in this report. Although computer scientists, ethics scholars and the media often refer to the notion of 'algorithmic bias', the choice of the authors of this report to speak of 'algorithmic discrimination' can be understood from the context of gender equality and non-discrimination law. The final section of the chapter also examines the interactions

7 See Kullmann, M (2018), 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law' 34 *International Journal of Comparative Labour Law and Industrial Relations*; Kim, PT (2017), 'Data-Driven Discrimination at Work' 58 *William and Mary Law Review* 857. See also, e.g. Wachter, S (2020), 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' 35 *Berkeley Technology Law Journal* 19.

of equality law with data protection law, a crucially relevant regulatory area in the context of the problem of algorithmic discrimination.

Algorithmic discrimination: what are the challenges for EU gender equality and non-discrimination law?

The second chapter of this report builds on the insights offered in Chapter 1 and analyses risks of algorithmic discrimination in the context of the European Union legal framework for the protection against discrimination. The chapter examines to what extent EU equality law is apposite to effectively capture and redress algorithmic discrimination and analyses the gaps and weaknesses of the legal protection in place. It argues that EU gender equality and non-discrimination law, while it offers important safeguards, also displays a number of inconsistencies, ambiguities and shortcomings that limit its ability to capture algorithmic discrimination in its various forms.

As regards the scope of EU equality law, the lack of protection against discrimination based on age, disability, sexual orientation and religion or beliefs in the area of goods and services represents a massive gap in an era where algorithms are increasingly used to profile and target people for selling purposes. Although the protection against discrimination based on race or ethnic origin is comprehensive, exceptions to the scope of the Gender Goods and Services Directive (2004/113/EC) in relation to the media, advertising and education are highly problematic. These gaps in the legal framework weaken the extent to which EU equality law captures discriminatory risks arising from the use of algorithms.

In addition, the specific and changing types of discrimination produced by algorithms demand an adaptation of the conceptual and doctrinal categories of EU non-discrimination. Proxy discrimination questions the boundaries of the exhaustive list of protected grounds defined in Article 19 TFEU and sheds new light on the role and place of the non-exhaustive list of protected grounds to be found on Article 21 of the EU Charter of Fundamental Rights. Algorithmic profiling based on granular analysis of personal and behavioural data entails heightened risks of intersectional discrimination, a type of discrimination that the Court of Justice has so far failed to adequately recognise.⁸

Furthermore, algorithmic discrimination challenges the standard doctrinal paradigms of EU and national non-discrimination law and in particular blurs the frontiers between direct and indirect discrimination. In light of the difficulties in tracking differential treatment based on protected grounds in ‘black box’ algorithms, the notion of indirect discrimination might become a conceptual ‘refuge’ to capture the discriminatory wrongs of algorithms. This development might reduce legal certainty if it leads, by default, to the generalisation of the open-ended objective justification test applicable in indirect discrimination cases as opposed to the narrower pool of justifications available in direct discrimination cases. Chapter 2 also shows how questions of evidence, responsibility and enforcement are complicated in the context of the transparency and explainability challenge outlined above. All in all, algorithmic discrimination shines a new light on many of the ‘traditional’ problems and critiques of EU gender equality and non-discrimination law. This report argues that these issues must be taken seriously – now more than ever in light of the ‘scale and speed challenge’ highlighted above – if EU regulators and policymakers are to ensure effective legal protection against algorithmic discrimination.

The legal challenges of algorithmic discrimination in European countries

Building on the observation that most examples of algorithmic discrimination and large strands of the scholarly literature on the subject refer to the US context, this report dedicates a third chapter to the specific challenges of algorithmic discrimination that arise at national level in Europe. Chapter 3 opens with an extensive review of the use of algorithms in the public and the private sector in 31 countries (all EU-27 Member States, EEA countries and the United Kingdom). In Europe, algorithms are used by public authorities in various areas, including labour market policies (e.g. to profile job seekers and allocate

8 Judgment of 24 November 2016, *David L. Parris v. Trinity College Dublin and Others*, C-443/15, EU:C:2016:897.

resources), social welfare (e.g. to predict risks of social marginalisation), education (e.g. to rank, select and assign candidates to higher education institutions), policing and fraud detection (e.g. to detect tax evasion and fraud and to predict risks of crime), the administration of justice (e.g. to support or predict judicial decision making) and the regulation of media platforms (e.g. to identify and curb hate speech online). In the European private sector, algorithms are used in fields such as employment and platform work (e.g. in human resources recruitment processes or in the allocation of work on online platforms), banking and insurance (e.g. to predict credit risks and calculate insurance damages) and targeted advertising, price-setting and retail (e.g. to personalise prices and offers and to disseminate adverts).

Chapter 3 also maps out the specific problems that arise in relation to algorithmically supported decision making in these countries and offers concrete examples of these issues. National experts report six specific sets of discrimination issues in relation to the use of algorithms in their own countries, which largely reflect the characteristics and challenges of algorithms discussed in Chapter 1:

- 1) biases in data (e.g. the use of protected characteristics and structurally biased data in algorithms used to predict risks of unemployment);
- 2) the discriminatory effects of algorithms (e.g. the stigmatising effects of the over-surveillance of 'problem neighbourhoods' by fraud detection algorithms or discriminatory behavioural targeting and personal pricing);
- 3) transparency problems and lack of information (e.g. the impossibility for judges, civil society organisations, consumers or public authorities to access information about whether an algorithm is discriminatory);
- 4) difficulties in detecting and identify algorithmic discrimination (in particular the difficulties for users and potential victims to identify and prove even *prima facie* algorithmic discrimination);
- 5) responsibility issues (e.g. identifying which of the many individuals and bodies involved in the design and use of algorithms is responsible for discrimination as well as jurisdictional issues when such organisations are established globally);
- 6) the gender digital gap in Europe (that is the stark under-representation of women and minority groups in science, technology and engineering education and professions).

Chapter 3 then discusses how public opinion, scientific communities and policymakers address such issues at national level in the 31 countries. All in all, it appears that public awareness of, and public action against, algorithmic discrimination as a specific issue remains limited in Europe. When discussed, algorithmic discrimination is often only considered as part of a wider set of concerns pertaining to data protection and fundamental rights in the ambit of AI in general. National scholarship also tends to examine the question of discrimination from the angle of data protection rather than that of equality law, and specific legal literature on the challenges posed by algorithms to national non-discrimination law remains limited so far. Although some policy discussions exist in some European countries, the national experts report that, so far, no new legislation or legislative reforms have been adopted to counter problems of algorithmic discrimination. Evaluating the legal framework in place in their country, national experts report that although gender equality and non-discrimination law, data protection law as well as technology-specific legislation, sectoral legislation and general criminal and civil law provisions can play a role in tackling algorithmic discrimination, specific gaps exist that could lead to a problematic lack of remedies at national level. The gaps noted mirror the ones that have been identified above, and mainly relate to the limitations of scope and the inability of current legislation to deal with the specific characteristics of algorithmic discrimination. In a majority of European countries, national courts have not yet been confronted with gender inequality or discrimination caused by algorithms. When judges have pondered cases involving algorithms, these usually relate to different issues such as transparency and data-protection. Despite the limited amount of litigation, cases concerning algorithmic discrimination are pending before some national courts at the time of writing and these developments will need to be followed closely.

Enforcing algorithmic equality: solutions and opportunities for gender equality and non-discrimination

Although the report highlights numerous challenges, risks and problems linked to the use of algorithms when it comes to ensuring equality and combating discrimination, the last chapter closes with a more positive outlook. Chapter 4 begins by highlighting a number of benefits and opportunities for equality and non-discrimination that the increasing use of algorithms makes possible. In particular, and by contrast to the human brain, algorithms offer opportunities to better visualise, measure, detect and ultimately correct discriminatory biases if proper legal regulation and public policy is put in place. For example, algorithms offer new opportunities to detect discriminatory job adverts on a large scale as well as to improve gender equality in recruitment processes.

In contrast to the absence of specific legislative reform covered in Chapter 3, Chapter 4 offers insights into a wide range of public and private good practice at national level to monitor and address algorithmic discrimination and to diversify the relevant professional communities. This includes the creation of dedicated monitoring and supervising institutions, the creation of soft-law instruments such as ethical codes, self-regulation practices such as voluntary codes of conduct, the publication of recommendations and guidelines, cooperation between data protection agencies and equality bodies and the setting up of public-private alliances. Various public policies and private initiatives have also been set up in Europe to diversify IT-related education and professions, which is also an important component of creating non-discriminatory AI. However, such measures overwhelmingly only promote gender equality as opposed to a more encompassing diversification of IT-related education and professions. Indeed, a better representation of all minority groups in these fields would favour a diversity of perspectives, which is a key first step towards algorithmic equality.

The report concludes by proposing a new integrated framework that offers a set of legal, educational and knowledge-based and technological measures and solutions to prevent, address and redress algorithmic discrimination.

At the legal level, such measures notably include adopting the draft Horizontal Directive under negotiation at the Council since 2008 to equalise the scope of EU non-discrimination law, addressing the gaps linked to the exceptions in the material scope of the Gender Goods and Services Directive in relation to the media, advertising and education, and bringing clarity on the prohibition of intersectional discrimination. Turning to the role of the Court of Justice, an expansive interpretation of the personal scope of EU equality law – both in terms of the scope of single protected grounds and the exhaustive nature of the list established in Article 19 TFEU in light of the open-ended provision in Article 21 of the Charter – would enhance the law's capacity to address algorithmic discrimination. Moreover, it is proposed that the concept of 'instruction to discriminate' be used as a doctrinal complement to direct and indirect discrimination in order to offer a better conceptual fit for algorithmic discrimination and, in parallel, reduce procedural and enforcement problems.⁹ As regards institutional solutions, it is suggested that a public and collective enforcement mechanism, such as an EU equality body, could considerably ease the monitoring and redress of algorithmic discrimination as the current burden of litigating algorithmic discrimination for individual victims is disproportionately high in the context of the transparency and explainability challenge. Finally, it is argued that the EU could foster the creation of an accreditation system for certification and supervision in relation to algorithmic discrimination. Such a certification ecosystem could promote an 'equality by design' approach to building algorithms from the get-go, promote the use of non-discriminatory algorithms by private and public bodies and facilitate redress by increasing transparency and explainability.

At the level of knowledge-based solutions, raising awareness about the risks of algorithmic discrimination among regulators, judges, economic players, the IT sector, and the society at large is crucial. Funding

⁹ On this argument, see Xenidis, R, 'Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law' (on file with the author).

research on algorithmic discrimination will also be key to informed practices in the design, use and regulation of algorithms. The monitoring of algorithmic discrimination should be supported by reporting tools, which could encourage watchdogs and whistleblowers to signal suspicions of algorithmic discrimination, which in turn would draw public attention to problematic practices. Finally, training and education is essential to preventing and combating algorithmic discrimination: by analogy to the ethical training of medical staff, not only IT specialists but also all relevant professional communities including regulators, judges, equality bodies, and so on should be educated and trained in the risks of algorithmic discrimination and ways to tackle it.

Finally, technological solutions include implementing preventive strategies (equality impact assessments and equality by design strategies) in the design, training and development phases of the creation of algorithms. In particular, various technological debiasing strategies have been developed by computer scientists to minimise algorithmic discrimination both at the level of data selection, labelling and use, and at the level of algorithmic models themselves. Technological solutions can also intervene *ex post*, in particular through the use of screening and auditing algorithms that can detect discrimination. Such technological instruments could prove extremely useful in the monitoring of algorithmic discrimination as well as in the ambit of the certification of algorithms as non-discriminatory. The development of such technological instruments should be encouraged by European public authorities.

These various legal, knowledge-based and technological solutions should be integrated in an interdisciplinary fashion. This report proposes the **PROTECT** framework as a set of key recommendations for public action in Europe. These recommendations revolve around seven key actions.

- ▶ **PREVENT**: through diverse and well-trained IT teams, equality impact assessments, *ex ante* 'equality by design' or 'legality by design' strategies.
- ▶ **REDRESS**: combine different legal tools in non-discrimination law, data protection law etc. to foster clear attribution of legal responsibilities, clear remedies, fair rules of evidence, flexible and responsive interpretation and application of non-discrimination concepts.
- ▶ **OPEN**: foster transparency, e.g. through open data requirements for monitoring purposes (such as access to source codes).
- ▶ **TRAIN**: educate, create and disseminate knowledge on non-discrimination and equality issues among IT specialists, raise awareness about issues of algorithmic discrimination with regulators, judges, recruiters, officials and society at large.
- ▶ **EXPLAIN**: establish explainability, accountability and information requirements;
- ▶ **CONTROL**: active human involvement (human-centred AI), e.g. in the form of human-in-the-loop (HITL) systems designed to avoid rubber-stamping, complemented by supervision and consultation mechanisms (chain of control and consultation with users).
- ▶ **TEST**: continuously monitor high-risk algorithms and their output, set up auditing, labelling and certification mechanisms.

Résumé

On assiste depuis quelques années à une prolifération de reportages médiatiques et d'articles scientifiques à propos de la discrimination algorithmique. On a notamment appris ainsi en 2019 que l'algorithme de la carte de paiement d'Apple accordait une limite de crédit plus élevée aux hommes qu'aux femmes alors que celles-ci ont de meilleures cotes de solvabilité;¹ et qu'Amazon a développé en 2014 un prototype de recrutement algorithmique dont le caractère discriminatoire envers les femmes a été établi par la suite et qui a dès lors dû être abandonné.² Une récente étude empirique montre à quel point le ciblage des annonces en ligne peut renforcer les stéréotypes et la ségrégation sur le marché du travail: lors de cette expérience, les chercheurs ont utilisé la plateforme Facebook pour diffuser de façon neutre diverses offres d'emploi. En définitive, les annonces concernant les postes de caissiers/caissières en supermarché ont touché une audience composée de 85% de femmes tandis que les offres d'emplois de chauffeurs de taxi ont touché une audience à 75% noire et celles pour des emplois de bûcherons une audience masculine à 90% et blanche à 72%.³ Il existe aujourd'hui un consensus global parmi les chercheurs comme parmi les décideurs pour considérer que les risques de discrimination algorithmique sont omniprésents et revêtent de multiples facettes. Garantir l'égalité et combattre la discrimination impose, dans ce contexte, de comprendre les risques en cause et les défis juridiques qui en découlent.

De fait, la Commission européenne a publié en 2019 un livre blanc sur l'intelligence artificielle qui reconnaît que l'utilisation croissante d'algorithmes en Europe engendre des risques spécifiques en termes de protection des droits fondamentaux, et en termes d'égalité et de non-discrimination en particulier.⁴ Ces risques sont également reconnus par la récente stratégie en faveur de l'égalité entre les hommes et les femmes 2020-2025 publiée par la Commission, qui admet que «l'IA [...] risque d'intensifier les inégalités entre les hommes et les femmes».⁵ En réponse, l'UE a demandé la création d'un «écosystème de confiance» exigeant que «l'IA européenne soit fondée sur nos valeurs et nos droits fondamentaux» – au cœur desquels figure le droit à l'égalité et à la non-discrimination.⁶

Le présent rapport analyse la manière dont la discrimination algorithmique met en question les garanties juridiques instaurées en Europe pour lutter contre la discrimination et assurer l'égalité de traitement. Il s'intéresse plus particulièrement à la capacité du cadre législatif relatif à l'égalité entre hommes et femmes et la non-discrimination actuellement en place au sein de l'UE à saisir la discrimination algorithmique et d'y remédier. Il étudie les lacunes et les faiblesses qui sont observées tant au niveau national qu'au niveau de l'UE par suite de l'interaction entre, d'une part, les formes spécifiques de discrimination qui se manifestent lorsque les systèmes décisionnels utilisent des algorithmes et, d'autre part, le champ d'application matériel et personnel particulier du cadre législatif existant. Le rapport recense également les solutions juridiques, les mesures d'accompagnement et les bonnes pratiques existantes visant à combattre la discrimination algorithmique et à y remédier, tant au niveau de l'Union qu'au niveau des

- 1 Gupta, AH (2019), «Are Algorithms Sexist?», *The New York Times* (15 novembre 2019), disponible sur: www.nytimes.com/2019/11/15/us/apple-card-goldman-sachs.html.
- 2 Dastin, J (2018), «Amazon scraps secret AI recruiting tool that showed bias against women» (10 octobre 2018), Reuters, disponible sur: www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.
- 3 Les auteurs précisent que ces chiffres représentent «les cas les plus extrêmes» de distribution biaisée. Dans l'expérience menée, ils ont sélectionné un public identique pour les trois annonces. Ali, M et al (2019), «Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes», service de préimpression arXiv disponible sur: [arXiv:1904.02095](https://arxiv.org/abs/1904.02095) 1.
- 4 Commission européenne (2020), «Livre blanc sur l'intelligence artificielle: une approche européenne axée sur l'excellence et la confiance», COM(2020) 65 final (Bruxelles 2020), p. 3, p. 11.
- 5 Commission européenne (2020), Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions «Une Union de l'égalité: stratégie en faveur de l'égalité entre les hommes et les femmes 2020-2025», COM(2020) 152 final (Bruxelles 2020), p. 6.
- 6 Commission européenne, «Livre blanc sur l'intelligence artificielle», p. 2.

différents pays. Il propose en outre son propre ensemble intégré de solutions juridiques, technologiques et fondées sur la connaissance en réponse au problème de la discrimination algorithmique.

Algorithmes et discrimination: de quoi s'agit-il?

Le premier chapitre fixe le cadre dans lequel va s'inscrire l'examen des diverses problématiques que suscitent les algorithmes par rapport au droit en matière d'égalité hommes-femmes et de non-discrimination. Il propose une série de définitions clés ainsi qu'une introduction aux divers usages possibles des algorithmes et aux différents types d'algorithmes actuellement utilisés. Étant donné que les défis pour le droit en matière d'égalité hommes-femmes et de non-discrimination varient selon les technologies algorithmiques utilisées, il est important de les différencier et d'identifier les problèmes qui y sont spécifiquement associés. La section 1.2 de ce premier chapitre, consacrée aux types d'algorithmes, s'adresse principalement aux lecteurs qui ne sont guère familiarisés avec les différentes technologies algorithmiques et leurs caractéristiques; les lecteurs qui en ont déjà connaissance peuvent donc sauter la section 1.2 pour se rendre directement à la section 1.3. Celle-ci décrit les diverses phases au cours desquelles une discrimination peut s'insinuer dans les algorithmes: de la conception à l'utilisation et de la programmation au développement et à la prise de décision, ceux-ci peuvent en effet être biaisés de plusieurs manières.

Le chapitre 1 explique également comment les préjugés humains et les stéréotypes de même que les inégalités sociétales structurelles, que reflètent les données utilisées pour entraîner les algorithmes, peuvent rendre ceux-ci discriminatoires. Rassemblant ces différents éléments de réflexion, la section principale du premier chapitre met en évidence six grands défis posés par les algorithmes au droit en matière d'égalité hommes-femmes et de non-discrimination:

- 1) *le défi lié au facteur humain, aux stéréotypes et aux biais cognitifs* décrit comment les partis pris implicites, les stéréotypes nuisibles et les préjugés discriminatoires véhiculés par l'homme risquent de contaminer les algorithmes créés par celui-ci, et de quelle manière les biais d'automatisation et d'ancrage aggravent ce risque;
- 2) *le défi des données* montre comment les données incarnent les formes historiquement consolidées de discrimination qui structurent la société, et comment le fait d'entraîner des algorithmes avec de telles données biaisées, ou avec des données inexactes, non représentatives ou déséquilibrées, conduit à une reproduction de ces inégalités structurelles par les algorithmes;
- 3) *le défi de la corrélation et des variables substitutives (proxies)*: en ce qui concerne le défi lié à la corrélation, le rapport explique ici comment les algorithmes peuvent réifier et perpétuer des corrélations discriminatoires en les considérant comme des relations causales et en les utilisant comme fondements pour de futures décisions, recommandations ou prévisions (ainsi par exemple une corrélation négative peut exister entre le genre et la performance au travail, non pas en raison d'un lien de causalité mais parce que les femmes ont été historiquement et systématiquement évaluées de manière plus négative que les hommes pour une même performance au travail)⁷; quant au défi lié aux variables substitutives, le rapport montre que la suppression de certaines caractéristiques protégées de la série de variables d'entrée disponibles s'avère insuffisante à la lumière de la capacité des algorithmes d'apprentissage à détecter des critères de substitution pour ces caractéristiques protégées;
- 4) *le défi de la transparence et de l'explicabilité* concerne la difficulté de surveiller et de prouver la discrimination algorithmique étant donné l'opacité de certains types d'algorithmes (même pour des informaticiens) et le manque d'information sur leur fonctionnement interne (en particulier lorsqu'il s'agit de codes et de données propriétaires);

7 Voir Kullmann, M (2018), «Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law», *International Journal of Comparative Labour Law and Industrial Relations*, vol.34; Kim, PT (2017), «Data-Driven Discrimination at Work» *William and Mary Law Review*, vol.58, p.857. Voir également Wachter, S (2020), «Affinity Profiling and Discrimination by Association in Online Behavioural Advertising», *Berkeley Technology Law Journal*, vol.35, p. 19.

- 5) *le défi de l'échelle et de la vitesse* concerne la manière dont la discrimination algorithmique peut «se propager» à plus grande échelle et beaucoup plus rapidement que la discrimination «humaine» du fait que les algorithmes accroissent à la fois la vitesse et l'échelle des processus décisionnels; et
- 6) *le défi de la responsabilité, de l'imputabilité et de la redevabilité* réside pour sa part dans la difficulté d'identifier qui doit être tenu pour responsable d'une issue discriminatoire dans le cadre de relations homme-machine extrêmement complexes, étant donné le nombre élevé d'intervenants différents impliqués dans la conception, la commercialisation et l'utilisation d'algorithmes.

Le chapitre 1 s'achève par un bref examen de la terminologie adoptée par le rapport. Bien que les informaticiens, les spécialistes en matière d'éthique et les médias fassent souvent référence à la notion de «biais algorithmique» (*algorithmic bias*), le parti pris par les auteurs du présent rapport de parler de «discrimination algorithmique» s'explique dans la perspective du droit en matière d'égalité entre les hommes et les femmes et de non-discrimination. La dernière section du premier chapitre se penche également sur les interactions du droit relatif à l'égalité avec le droit en matière de protection des données, domaine réglementaire qui revêt une pertinence toute particulière face au problème de la discrimination algorithmique.

Discrimination algorithmique: quels sont les défis pour le droit de l'UE relatif à l'égalité hommes-femmes et à la non-discrimination?

Le deuxième chapitre de ce rapport développe les réflexions proposées au chapitre 1 et procède à une analyse des risques de discrimination algorithmique s'inscrivant dans le cadre juridique antidiscriminatoire adopté par l'Union européenne. Il étudie dans quelle mesure le droit européen en matière d'égalité est pertinent pour saisir effectivement la discrimination algorithmique et y remédier, et se penche sur les lacunes et faiblesses de la protection juridique en place. Il fait valoir que, tout en offrant d'importantes garanties, le droit de l'UE en matière d'égalité hommes-femmes et de non-discrimination n'en présente pas moins une série d'incohérences, d'ambiguïtés et de carences qui en limitent la capacité à saisir la discrimination algorithmique sous ses différentes formes.

En ce qui concerne le champ d'application de ce droit de l'UE, l'absence de protection contre la discrimination fondée sur l'âge, le handicap, l'orientation sexuelle et la religion ou les convictions dans le domaine des biens et des services constitue une lacune majeure à l'heure où les algorithmes sont de plus en plus souvent utilisés pour établir le profil et cibler des individus à des fins commerciales. La protection contre la discrimination fondée sur la race ou l'origine ethnique est très étendue, mais les exceptions au champ d'application de la directive relative à l'égalité de traitement entre hommes et femmes dans l'accès et la fourniture de biens et de services (2004/113/CE) pour ce qui concerne les médias, la publicité et l'éducation sont très préoccupantes. Ces lacunes au niveau du cadre légal affaiblissent la capacité du droit de l'UE en matière d'égalité à saisir les risques discriminatoires découlant de l'usage d'algorithmes.

De surcroît, les formes particulières et changeantes de discrimination produites par les algorithmes requièrent l'adaptation des catégories conceptuelles et doctrinales de la non-discrimination au niveau de l'UE. La discrimination fondée sur un critère de substitution (*proxy discrimination*) met en question les limites de la liste exhaustive de motifs protégés définis à l'article 19 du TFUE et apporte un nouvel éclairage sur le rôle et la place de la liste non exhaustive de motifs protégés qui figure à l'article 21 de la Charte des droits fondamentaux de l'Union européenne. Le profilage algorithmique fondé sur l'analyse détaillée des données personnelles et comportementales entraîne un risque accru de discrimination intersectionnelle – forme de discrimination que la Cour de justice n'a pas encore reconnue de façon adéquate à ce jour.⁸

La discrimination algorithmique remet en outre en question les paradigmes doctrinaux classiques du droit antidiscriminatoire national et de l'UE, et brouille les frontières entre discrimination directe et

8 Arrêt du 24 novembre 2016, *David L. Parris c. Trinity College Dublin e.a.*, C-443/15, EU:C:2016:897.

discrimination indirecte. Au vu des difficultés posées par le repérage d'une différence de traitement basée sur un motif protégé dans des algorithmes fonctionnant en «boîtes noires», la notion de discrimination indirecte pourrait bien devenir un «refuge» conceptuel pour saisir les injustices discriminatoires des algorithmes. Une telle évolution amoindrirait la sécurité juridique si elle devait déboucher, par défaut, sur une généralisation du test ouvert fondé sur la justification objective applicable dans les cas de discrimination indirecte, par opposition à l'éventail plus étroit des justifications disponibles lorsqu'il s'agit de discrimination directe. Le chapitre 2 montre également que les questions de preuve, de responsabilité et de mise en application sont particulièrement complexes au vu du défi de transparence et d'explicabilité décrit plus haut. Au total, la discrimination algorithmique apporte un nouvel éclairage sur bon nombre de critiques et de problèmes «traditionnels» concernant le droit européen relatif à l'égalité hommes-femmes et à la lutte contre la discrimination. Le présent rapport soutient que ces questions doivent être prises au sérieux – aujourd'hui plus que jamais face au «défi de l'échelle et de la vitesse» évoqué ci-dessus – pour que les régulateurs et décideurs de l'UE parviennent à assurer une protection juridique efficace contre la discrimination algorithmique.

Les défis juridiques posés par la discrimination algorithmique dans les pays européens

Partant de l'observation selon laquelle la plupart des exemples de discrimination algorithmique et une large part des ouvrages de recherche sur ce thème font référence au contexte américain, le rapport consacre son troisième chapitre aux défis liés à la discrimination algorithmique qui se posent spécifiquement au niveau national en Europe. Le chapitre 3 commence par proposer un examen approfondi de l'utilisation des algorithmes dans le secteur public et le secteur privé de 31 pays (les pays membres de l'UE-27, les pays de l'EEE et le Royaume-Uni). En Europe, les algorithmes sont utilisés par les pouvoirs publics dans divers secteurs tels que les politiques liées au marché du travail (notamment pour l'établissement du profil des demandeurs d'emploi et l'allocation des ressources), à la sécurité sociale (prévision des risques de marginalisation sociale, par exemple), à l'éducation (classement, sélection et affectation des candidats au niveau des établissements d'enseignement supérieur, entre autres), à la surveillance et à la détection de la fraude (détection de l'évasion et de la fraude fiscales et prévision des risques de criminalité, par exemple), à l'administration de la justice (soutien ou prévision en matière de décisions judiciaires notamment) et à la réglementation des plateformes médiatiques (reconnaissance et lutte contre les discours haineux en ligne). Dans le secteur privé européen, les algorithmes sont utilisés dans des domaines tels que l'emploi et les plateformes de travail (notamment dans le cadre des processus de recrutement de ressources humaines ou dans l'attribution de travail sur les plateformes en ligne), la banque et l'assurance (prévision des risques de crédit et calcul des dommages d'assurance, par exemple) et le ciblage de la publicité, la fixation des prix et la promotion commerciale (personnalisation des prix et des offres et diffusion des annonces, entre autres).

Le chapitre 3 recense également les problèmes particuliers qui surviennent dans ces pays en raison de prises de décisions fondées sur les algorithmes, et il en propose des exemples concrets. Les experts nationaux font état de six séries spécifiques de problèmes de discrimination liée à l'usage d'algorithmes dans leurs pays respectifs, lesquelles reflètent largement les caractéristiques des algorithmes et leurs enjeux examinés au chapitre 1:

- 1) les biais des données (notamment l'usage de caractéristiques protégées et de données structurellement biaisées dans les algorithmes servant à prévoir les risques de chômage);
- 2) les effets discriminatoires des algorithmes (tels que les effets stigmatisants d'une surveillance excessive des «quartiers à problèmes» par des algorithmes de détection de la fraude, ou le caractère discriminatoire d'un ciblage comportemental ou d'une tarification individualisée);
- 3) les problèmes de transparence et le manque d'information (notamment l'impossibilité pour les magistrats, les organisations de la société civile, les consommateurs ou les autorités publiques d'accéder à des informations quant à la nature éventuellement discriminatoire d'un algorithme);

- 4) la difficulté de déceler et d'identifier une discrimination algorithmique (et en particulier la difficulté pour les utilisateurs et les victimes potentielles de repérer et de démontrer l'existence – voire même une présomption – de discrimination algorithmique);
- 5) la problématique de la responsabilité (comment, par exemple, déterminer qui, parmi les nombreuses personnes et organisations intervenant dans la conception et l'utilisation des algorithmes, est responsable de la discrimination; sans compter que des questions de compétence se posent également lorsque ces structures sont établies à l'échelle mondiale);
- 6) la fracture numérique hommes-femmes en Europe (autrement dit la sous-représentation flagrante des femmes et des groupes minoritaires dans les formations et professions relevant des sciences, des technologies, de l'ingénierie et des mathématiques).

Le chapitre 3 s'intéresse ensuite à la manière dont l'opinion publique, les communautés scientifiques et les décideurs des 31 pays abordent ces questions à l'échelon national. Il semble qu'au total la sensibilisation du grand public européen à l'égard de la discrimination algorithmique en tant que problématique spécifique, de même que l'action publique à son encontre, demeurent limitées. Lorsqu'il y a débat à son sujet, la discrimination algorithmique est souvent considérée comme une préoccupation parmi bien d'autres concernant la protection des données et les droits fondamentaux dans le cadre plus général de l'IA. Les spécialistes nationaux ont également tendance à envisager la question de la discrimination sous l'angle de la protection des données plutôt que sous celui du droit en matière d'égalité, et les ouvrages juridiques spécifiquement consacrés aux défis posés par les algorithmes au niveau du droit anti-discrimination national restent peu nombreux à ce jour. Des discussions politiques se tiennent actuellement dans plusieurs pays européens mais, selon les experts nationaux, aucune nouvelle législation ni réforme législative n'a été adoptée jusqu'ici pour résoudre les problèmes de discrimination algorithmique. L'évaluation du cadre juridique en place dans leurs pays respectifs conduit les experts nationaux à constater que si les dispositions du droit en matière d'égalité hommes-femmes et de non-discrimination, du droit en matière de protection des données ainsi que celles de la législation technospécifique, de la législation sectorielle et du droit civil et pénal général peuvent jouer un rôle dans la lutte contre la discrimination algorithmique, des lacunes spécifiques persistent néanmoins et sont susceptibles de générer un manque problématique de recours à l'échelon national. Les lacunes recensées reflètent celles identifiées plus haut, et concernent principalement les limites du champ d'application et l'incapacité de la législation actuelle à répondre aux spécificités de la discrimination algorithmique. Dans une majorité de pays européens, les juridictions nationales n'ont pas encore été confrontées à une inégalité de genre ou une discrimination causée par des algorithmes. Lorsque des magistrats ont été amenés à étudier des affaires impliquant des algorithmes, il s'agissait généralement de cas portant sur d'autres problématiques telles que la transparence et la protection des données. En dépit d'un contentieux peu abondant, des affaires relatives à la discrimination algorithmique sont en instance devant certaines juridictions nationales au moment de la rédaction de ce rapport et il conviendra d'en suivre les développements avec la plus grande attention.

Mise en application de l'égalité algorithmique: des solutions et des opportunités pour l'égalité hommes-femmes et la non-discrimination

Si le rapport met en évidence les multiples enjeux, risques et problèmes associés à l'utilisation d'algorithmes lorsqu'il s'agit de garantir l'égalité et de combattre la discrimination, son dernier chapitre se clôture cependant sur une note plus positive. Le chapitre 4 commence en effet par mettre en lumière une série d'avantages et d'opportunités rendus possibles par l'usage plus intensif d'algorithmes. Ainsi notamment, et contrairement au cerveau humain, les algorithmes offrent la possibilité de mieux visualiser, mesurer, déceler et, en définitive, remédier à des biais discriminatoires pour autant qu'une réglementation et une politique publique adéquates soient mises en place. Les algorithmes offrent par exemple de nouvelles possibilités de détecter des offres d'emploi discriminatoires à grande échelle, ou bien d'améliorer l'égalité entre hommes et femmes dans les processus de recrutement.

En dépit de l'absence de réforme législative spécifique abordée au chapitre 3, le chapitre 4 montre qu'il existe un large éventail de bonnes pratiques adoptées au niveau national dans les secteurs public

et privé pour surveiller et combattre la discrimination algorithmique et diversifier les communautés professionnelles concernées. Ces pratiques consistent notamment à créer des institutions spécialisées de surveillance et de contrôle, ainsi que des instruments non contraignants tels que des codes de déontologie, des pratiques d'autorégulation par exemple fondées sur des codes de conduite volontaires, la publication de recommandations et de lignes directrices, une coopération entre organismes en charge de la protection des données et organismes pour la promotion de l'égalité, et l'instauration d'alliances public-privé. Diverses politiques publiques et initiatives privées ont également vu le jour en Europe dans le but de diversifier davantage l'enseignement et les métiers liés aux technologies de l'information – ce qui devrait également contribuer à la création d'une IA non discriminatoire. Il est regrettable toutefois que la quasi-totalité de ces mesures visent à promouvoir uniquement l'égalité hommes-femmes au lieu d'ambitionner une diversification plus large des formations et professions liées aux technologies de l'information. Une meilleure représentation de tous les groupes minoritaires dans ces domaines favoriserait pourtant une plus grande diversité de points de vue, une première étape déterminante sur la voie de l'égalité algorithmique.

Le rapport propose en conclusion un cadre intégré novateur offrant une série de mesures et de solutions juridiques, technologiques et fondées sur l'éducation et les connaissances pour prévenir, traiter et remédier à la discrimination algorithmique.

Sur le plan juridique, ces mesures consistent notamment à adopter le projet de directive horizontale qui fait l'objet de négociations au niveau du Conseil depuis 2008 et qui est destinée à uniformiser le champ d'application du droit de l'UE en matière de non-discrimination; à combler les lacunes liées aux exceptions au champ d'application matériel de la directive relative à l'égalité de traitement entre hommes et femmes dans l'accès à des biens et des services pour ce qui concerne les médias, la publicité et l'éducation; et à préciser davantage l'interdiction de discrimination intersectionnelle. Pour ce qui est du rôle de la Cour de justice, une interprétation large du champ d'application personnel du droit de l'UE en matière d'égalité – en ce qui concerne à la fois le champ d'application des motifs protégés et l'exhaustivité de la liste dressée à l'article 19 TFUE à la lumière de la disposition ouverte de l'article 21 de la Charte – renforcerait la capacité de ce cadre juridique à aborder la discrimination algorithmique. Il est suggéré en outre d'utiliser le concept de «l'injonction de discriminer» en tant que complément doctrinal à la discrimination directe et indirecte en vue d'assurer une meilleure concordance conceptuelle avec la discrimination algorithmique et de limiter, dans le même temps, les problèmes de procédure et de mise en application.⁹ En ce qui concerne les solutions institutionnelles, le rapport suggère qu'un mécanisme public et collectif de mise en application, tel qu'un organisme européen de promotion de l'égalité, faciliterait considérablement le contrôle et l'élimination de la discrimination algorithmique car le déclenchement d'un contentieux représente actuellement pour des victimes individuelles une charge disproportionnellement élevée en raison des défis liés à la transparence et l'explicabilité. Le rapport avance enfin que l'UE pourrait encourager la création d'un système d'accréditation des instances de certification et de supervision pour ce qui concerne la discrimination algorithmique. Cet écosystème de certification pourrait favoriser une approche de la construction d'algorithmes qui soit axée sur l'égalité dès le stade de leur conception (*equality by design*); promouvoir l'utilisation d'algorithmes non discriminatoires par les organismes publics et privés; et faciliter les recours juridiques à travers un renforcement de la transparence et de l'explicabilité.

Pour ce qui concerne les solutions fondées sur les connaissances, il est essentiel de sensibiliser davantage les régulateurs, les magistrats, les acteurs économiques, le secteur des technologies de l'information et la société en général aux risques de discrimination algorithmique. Le financement de la recherche sur ce thème sera également déterminant pour étayer les pratiques au niveau de la conception, de l'utilisation et de la réglementation des algorithmes. La surveillance de la discrimination algorithmique devrait pouvoir s'appuyer sur certains outils de signalement, lesquels inciteraient les observateurs critiques et les lanceurs d'alerte à signaler leurs soupçons de discrimination algorithmique – démarche

9 Voir, à propos de cet argument, Xenidis, R, «Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law» (disponible auprès de l'auteur).

qui attirerait à son tour l'attention du public sur certaines pratiques problématiques. Enfin, la formation et l'éducation s'avèrent indispensables pour prévenir et combattre cette forme de discrimination: par analogie à la formation éthique du personnel médical, les spécialistes des technologies de l'information et les communautés professionnelles concernées, y compris les régulateurs, les magistrats, les organismes de promotion de l'égalité, etc., devraient être formés aux risques de la discrimination algorithmique et aux moyens d'y remédier.

Pour terminer, des solutions techniques prévoient la mise en œuvre de stratégies préventives (évaluations de l'impact des algorithmes sur l'égalité (*impact assessments*) et stratégies de développement intégrant l'égalité dès la conception des algorithmes (*equality by design*)) aux différents stades de la création des algorithmes (conception, entraînement et développement). Diverses stratégies technologiques de suppression des biais ont ainsi été plus particulièrement mises au point par des informaticiens en vue de minimiser la discrimination algorithmique, à la fois au niveau de la sélection des données, de leur étiquetage et de leur utilisation et au niveau des modèles algorithmiques eux-mêmes. Des solutions technologiques peuvent également être appliquées a posteriori: le rapport suggère notamment l'utilisation d'algorithmes de dépistage et de vérification capables de déceler les discriminations. De tels outils technologiques pourraient s'avérer extrêmement utiles pour la surveillance de la discrimination algorithmique ainsi que dans le cadre de la certification du caractère non discriminatoire des algorithmes. Les autorités publiques européennes devraient encourager le développement d'instruments technologiques de ce type.

Il convient de procéder à une intégration interdisciplinaire de ces diverses solutions juridiques, technologiques et fondées sur les connaissances, et le présent rapport propose la structure **PROTECT** pour regrouper les principales recommandations d'action publique en Europe. Ces recommandations s'articulent autour de sept actions clés:

- ▶ **PRÉVENIR**: faire réaliser préalablement, par des équipes diversifiées et dûment formées, des études d'impact sur l'égalité ainsi que des stratégies axées sur l'égalité dès la conception (*equality by design*) et la légalité dès la conception (*legality by design*).
- ▶ **RÉPARER**: combiner différents instruments juridiques du droit antidiscriminatoire, du droit relatif à la protection des données, etc. afin d'instaurer une attribution explicite des responsabilités légales, des recours clairs, des règles équitables en matière de preuve, une interprétation et une mise en application flexible et réactive des concepts du droit antidiscriminatoire.
- ▶ **OUVRIRE**: favoriser la transparence au moyen notamment d'obligations en matière d'accès libre aux données à des fins de contrôle (accès aux codes sources, par exemple);
- ▶ **TRAVAILLER À FORMER**: éduquer, créer et diffuser les connaissances relatives aux questions de non-discrimination et d'égalité parmi les spécialistes des technologies de l'information, et sensibiliser les régulateurs, les magistrats, les recruteurs, les fonctionnaires et la société en général à la problématique de la discrimination algorithmique.
- ▶ **EXPLIQUER**: définir des exigences en matière d'explicabilité, de responsabilité et d'information.
- ▶ **CONTRÔLER**: veiller à une implication humaine active (IA centrée sur l'homme), en recourant par exemple à des systèmes HITL (*Human in the Loop*) conçus pour éviter la validation automatique, et accompagnée de mécanismes de supervision et de consultation (chaîne de contrôle et consultation des usagers concernés).
- ▶ **TESTER**: surveiller en permanence les algorithmes à haut risque et leurs résultats, et mettre en place des mécanismes d'audit, de labellisation et de certification.

Zusammenfassung

In den letzten Jahren hat es eine Vielzahl von Medienberichten und wissenschaftlichen Artikeln zum Thema algorithmische Diskriminierung gegeben. 2019 wurde zum Beispiel bekannt, dass der Algorithmus der Apple Card Männern ein höheres Kreditlimit einräumte als Frauen, obwohl letztere eine bessere Bonität haben.¹ 2014 entwickelte Amazon einen Prototyp für ein algorithmisches Recruiting Tool, der sich später als diskriminierend für Frauen herausstellte und eingestellt werden musste.² Eine aktuelle empirische Studie hat gezeigt, wie das Targeting von Online-Anzeigen Stereotypen und Segregation auf dem Arbeitsmarkt verstärken kann: Für ihr Experiment benutzten die Forscher die Anzeigenplattform von Facebook, um verschiedene Stellenanzeigen neutral zu verbreiten; letztendlich erreichten Stellenangebote für Kassierkräfte in Supermärkten ein Publikum, das zu 85 % weiblich war, Stellenangebote für Taxifahrer hingegen ein zu 75 % schwarzes Publikum und Stellenangebote für Holzfäller ein Publikum, das zu 90 % männlich und zu 72 % weiß war.³ Unter Forschenden und politischen Entscheidungsträgern besteht allgemeiner Konsens darüber, dass die Gefahren algorithmischer Diskriminierung allgegenwärtig und vielschichtig sind. Um Gleichheit zu gewährleisten und Diskriminierung zu bekämpfen, ist es in diesem Zusammenhang von grundlegender Bedeutung, diese Gefahren und die damit verbundenen rechtlichen Herausforderungen zu verstehen.

Tatsächlich hat die Europäische Kommission 2019 ein Weißbuch zu künstlicher Intelligenz veröffentlicht, in dem festgestellt wurde, dass der zunehmende Einsatz von Algorithmen in Europa spezifische Risiken in Bezug auf den Schutz der Grundrechte und insbesondere in Bezug auf Gleichheit und Nichtdiskriminierung birgt.⁴ Diese Risiken werden auch in der jüngst von der Kommission veröffentlichten „Strategie für die Gleichheit der Geschlechter 2020-2025“ konstatiert, der zufolge „KI [...] Ungleichheiten zwischen Frauen und Männern verstärken [kann]“.⁵ Angesichts dessen hat die EU dazu aufgerufen, ein „Ökosystem für Vertrauen“ zu schaffen; Voraussetzung dafür ist, dass „die europäische KI auf unseren [der EU] Werten und Grundrechten ... fußt“, wobei dem Recht auf Gleichheit und Nichtdiskriminierung zentrale Bedeutung zukommt.⁶

Der Bericht untersucht, wie algorithmische Diskriminierung die rechtlichen Garantien herausfordert, die in Europa geschaffen wurden, um Diskriminierung zu bekämpfen und Gleichheit zu gewährleisten. Dabei geht er insbesondere der Frage nach, ob und wie der gegenwärtige EU-Rechtsrahmen für Geschlechtergleichstellung und Nichtdiskriminierung in der Lage ist, algorithmische Diskriminierung angemessen zu erfassen und zu beseitigen. Er erkundet die Lücken und Schwächen, die sowohl auf nationaler als auch auf EU-Ebene aus der Wechselwirkung zwischen den spezifischen Formen von Diskriminierung, die beim Einsatz von Algorithmen in Entscheidungssystemen auftreten, einerseits und dem speziellen sachlichen und persönlichen Geltungsbereich des bestehenden Rechtsrahmens andererseits entstehen. Außerdem zeigt der Bericht auf, welche rechtlichen Lösungen, begleitenden politischen Maßnahmen und bewährten Verfahren existieren, um sowohl auf nationaler als auch auf EU-

- 1 Gupta, A. H. (2019), „Are Algorithms Sexist?“, *The New York Times* (15. November 2019), abrufbar unter: www.nytimes.com/2019/11/15/us/apple-card-goldman-sachs.html.
- 2 Dastin, J. (2018), „Amazon scraps secret AI recruiting tool that showed bias against women“, (10. Oktober 2018), *Reuters*, abrufbar unter: www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.
- 3 Die Autoren weisen darauf hin, dass diese Zahlen „die extremsten Fälle“ einer ungleichen Verteilung darstellen. In dem Experiment wählten sie für alle drei Anzeigen ein identisches Publikum. Ali, M. u. a. (2019), „Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes“, arXiv Preprint, abrufbar unter: arXiv:1904.02095 1.
- 4 Europäische Kommission (2020), Weißbuch zur Künstlichen Intelligenz — ein europäisches Konzept für Exzellenz und Vertrauen, COM(2020) 65 final (Brüssel 2020), S. 3.
- 5 Europäische Kommission (2020), Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen „Eine Union der Gleichheit: Strategie für die Gleichheit der Geschlechter 2020-2025“, COM(2020) 152 final (Brüssel 2020), S. 7.
- 6 Europäische Kommission (2020), „Weißbuch zur Künstlichen Intelligenz“, S. 2.

Ebene gegen algorithmische Diskriminierung vorzugehen und diese zu beseitigen. Schließlich formuliert der Bericht eine Reihe integrierter rechtlicher, wissenschaftlicher und technologischer Lösungsvorschläge für das Problem der algorithmischen Diskriminierung.

Algorithmen und Diskriminierung: Worum geht es genau?

Das erste Kapitel des Berichts steckt den Rahmen für die Untersuchung der verschiedenen Fragen ab, die Algorithmen in Bezug auf das Gleichstellungs- und Antidiskriminierungsrecht aufwerfen. Es liefert eine Reihe von Schlüsseldefinitionen sowie eine Einführung in die verschiedenen Einsatzmöglichkeiten von Algorithmen und in die verschiedenen Arten von Algorithmen, die derzeit zum Einsatz kommen. Da unterschiedliche algorithmische Technologien das Gleichstellungs- und Antidiskriminierungsrecht vor unterschiedliche Herausforderungen stellen, ist es wichtig, zwischen den verschiedenen Arten und den spezifischen Problemen, die sie aufwerfen, zu unterscheiden. Abschnitt 1.2 über Arten von Algorithmen richtet sich in erster Linie an Leserinnen und Leser, die mit den verschiedenen algorithmischen Technologien und ihren Merkmalen nicht vertraut sind. Leserinnen und Leser, die mit diesen bereits vertraut sind, können Abschnitt 1.2 überspringen und direkt mit Abschnitt 1.3 fortfahren. Letzterer beschreibt die verschiedenen Phasen, in denen sich Diskriminierung in Algorithmen einschleichen kann: Vom Entwurf bis zur Anwendung und von der Planung über die Entwicklung bis zur Entscheidungsfindung können Verzerrungseffekte Algorithmen auf unterschiedliche Weise beeinflussen.

In Kapitel 1 wird auch erläutert, wie menschliche Vorurteile und Stereotype sowie strukturelle gesellschaftliche Ungleichheiten, die sich in den zum Training von Algorithmen verwendeten Daten widerspiegeln, zu diskriminierenden Algorithmen führen können. Indem er diese verschiedenen Erkenntnisse zusammenführt, zeigt der zentrale Teil von Kapitel 1 sechs große Herausforderungen auf, vor die Algorithmen das Gleichstellungs- und Antidiskriminierungsrecht stellen:

- 1) die *Herausforderung des menschlichen Faktors, der Stereotypisierung und der kognitiven Verzerrung* beschreibt, wie implizite Verzerrungseffekte, schädliche Stereotype und diskriminierende Vorurteile, die Menschen mit sich herumtragen, Gefahr laufen, von Menschen gemachte Algorithmen zu „infizieren“, und wie Automatisierung und Ankereffekte diese Gefahr verstärken;
- 2) die *Herausforderung der Daten* beschreibt, wie Daten historisch verfestigte Diskriminierungsmuster enthalten, die die Gesellschaft strukturieren, und wie das Trainieren von Algorithmen mit solchen verzerrten Daten – bzw. mit fehlerhaften, nicht repräsentativen oder unausgewogenen Daten – dazu führt, dass diese Algorithmen strukturelle Ungleichheiten reproduzieren;
- 3) die *Herausforderung der Korrelationen und Proxys*: Die Herausforderung der Korrelationen beschreibt, wie Algorithmen diskriminierende Korrelationen reifizieren und perpetuieren können, indem sie diese als kausale Beziehungen betrachten und als Grundlage für künftige Entscheidungen, Empfehlungen und Prognosen heranziehen (z. B. kann eine negative Korrelation zwischen Geschlecht und Arbeitsleistung bestehen – nicht etwa weil ein kausaler Zusammenhang bestünde, sondern weil Frauen seit jeher bei gleicher Arbeitsleistung systematisch schlechter bewertet werden als Männer);⁷ die Herausforderung der Proxys beschreibt, warum das Entfernen geschützter Merkmale aus dem Pool verfügbarer Eingabevariablen angesichts der Fähigkeit von Lernalgorithmen, Proxys für diese geschützten Merkmale zu finden, nicht ausreicht;
- 4) die *Herausforderung der Transparenz und Erklärbarkeit* betrifft die Schwierigkeit, angesichts der (selbst für Informatiker bestehenden) Undurchsichtigkeit bestimmter Arten von Algorithmen und des Mangels an Informationen über ihre interne Funktionsweise (vor allem wenn Codes und Daten proprietär sind) algorithmische Diskriminierung zu überwachen und nachzuweisen;
- 5) die *Herausforderung des Umfangs und der Geschwindigkeit* beschreibt, wie algorithmische Diskriminierung sich in größerem Umfang und mit einer viel höheren Geschwindigkeit als

7 Vgl. Kullmann, M. (2018), „Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law“ 34 *International Journal of Comparative Labour Law and Industrial Relations*; Kim, PT (2017), „Data-Driven Discrimination at Work“ 58 *William & Mary Law Review* 857; vgl. ebenso Wachter, S. (2020), „Affinity Profiling and Discrimination by Association in Online Behavioural Advertising“ 35 *Berkeley Technology Law Journal* 19.

„menschliche“ Diskriminierung ausbreiten kann, da Algorithmen sowohl die Geschwindigkeit als auch den Umfang von Entscheidungsprozessen steigern;

- 6) die *Herausforderung der Verantwortung, Haftung und Rechenschaftspflicht* liegt in der Schwierigkeit zu bestimmen, wer – angesichts der Vielzahl unterschiedlicher Akteure, die an der Entwicklung, Vermarktung und Anwendung von Algorithmen beteiligt sind – in komplexen Mensch-Maschine-Beziehungen für ein diskriminierendes Ergebnis verantwortlich, haftbar und/oder rechenschaftspflichtig ist.

Abschließend geht Kapitel 1 kurz auf die in dem Bericht verwendete Terminologie ein. Wenngleich Informatiker, Ethikexperten und die Medien häufig den Begriff „algorithmisches Bias“ verwenden, ist die Entscheidung der Autorinnen dieses Berichts, von „algorithmischer Diskriminierung“ zu sprechen, im Kontext des Gleichstellungs- und Antidiskriminierungsrechts zu verstehen. Im letzten Abschnitt des ersten Kapitels wird auch das Zusammenspiel von Gleichstellungsrecht und Datenschutzrecht behandelt, ein Regelungsbereich, der für das Problem der algorithmischen Diskriminierung von größter Bedeutung ist.

Algorithmische Diskriminierung: Was sind die Herausforderungen für das EU-Gleichstellungs- und Antidiskriminierungsrecht?

Das zweite Kapitel des Berichts baut auf den Erkenntnissen aus Kapitel 1 auf und analysiert die Risiken algorithmischer Diskriminierung im Kontext des Rechtsrahmens der Europäischen Union für den Schutz vor Diskriminierung. Das Kapitel geht der Frage nach, inwieweit die europäischen Rechtsvorschriften über Gleichbehandlung geeignet sind, algorithmische Diskriminierung wirksam zu erfassen und gegen diese vorzugehen, und untersucht die Lücken und Schwächen des bestehenden Rechtsschutzes. Es wird gezeigt, dass die EU-Rechtsvorschriften zur Gleichstellung von Frauen und Männern und zur Nichtdiskriminierung zwar wichtige Garantien enthalten, gleichzeitig jedoch eine Reihe von Unstimmigkeiten, Mehrdeutigkeiten und Mängeln aufweisen, die ihre Fähigkeit einschränken, algorithmische Diskriminierung in ihren unterschiedlichen Formen zu erfassen.

Was den Geltungsbereich der europäischen Gleichbehandlungsvorschriften betrifft, so stellt der fehlende Schutz vor Diskriminierung aufgrund des Alters, einer Behinderung, der sexuellen Orientierung und der Religion oder Weltanschauung im Bereich Güter und Dienstleistungen – in einer Zeit, in der Algorithmen zunehmend dafür eingesetzt werden, Profile von Menschen zu erstellen und für kommerzielle Zwecke zu nutzen – ein gravierendes Manko dar. Während der Schutz vor Diskriminierung aus Gründen der „Rasse“ oder der ethnischen Herkunft sehr umfassend ist, sind die Ausnahmen vom Geltungsbereich der Richtlinie über die Gleichbehandlung von Männern und Frauen beim Zugang zu und bei der Versorgung mit Gütern und Dienstleistungen (2004/113/EG) in Bezug auf Medien, Werbung und Bildung sehr bedenklich. Diese Lücken im Rechtsrahmen schwächen die Fähigkeit des EU-Gleichbehandlungsrechts, Diskriminierungsrisiken, die sich aus dem Einsatz von Algorithmen ergeben, zu erfassen.

Hinzu kommt, dass die spezifischen und sich verändernden Formen von Diskriminierung, die durch Algorithmen erzeugt werden – Proxy-Diskriminierung, intersektionelle Diskriminierung und Fehlkategorisierung –, eine Anpassung der konzeptuellen und doktrinen Nichtdiskriminierungskategorien der EU erforderlich machen. Proxy-Diskriminierung stellt die Grenzen der abschließenden Auflistung geschützter Merkmale in Artikel 19 AEUV in Frage und wirft ein neues Licht auf die Rolle und Stellung der nicht abschließenden Auflistung geschützter Merkmale in Artikel 21 der EU-Grundrechtecharta. Algorithmisches Profiling, das auf einer granularen Analyse personen- und verhaltensbezogener Daten basiert, birgt ein erhöhtes Risiko intersektioneller Diskriminierung – eine Form von Diskriminierung, der seitens des Gerichtshofs bisher nicht angemessen Rechnung getragen wurde.⁸

Algorithmische Diskriminierung stellt darüber hinaus die traditionellen doktrinen Paradigmen des nationalen und des EU-Antidiskriminierungsrechts in Frage und verwischt insbesondere die

8 Urteil vom 24. November 2016, *David L. Parris / Trinity College Dublin u. a.*, C-443/15, EU:C:2016:897.

Grenzen zwischen unmittelbarer und mittelbarer Diskriminierung. Angesichts der Schwierigkeiten, auf geschützten Merkmalen basierende Ungleichbehandlungen in Blackbox-Algorithmen nachzuverfolgen, könnte das Konzept der mittelbaren Diskriminierung zu einer begrifflichen „Zufluchtsstätte“ werden, um diskriminierende Fehlentscheidungen von Algorithmen zu erfassen. Eine solche Entwicklung könnte die Rechtssicherheit beeinträchtigen, wenn sie standardmäßig zu einer Generalisierung der offenen sachlichen Rechtfertigungsprüfung führt, die in Fällen von mittelbarer Diskriminierung zur Anwendung kommt, im Gegensatz zu der engeren Palette von Rechtfertigungen, die in Fällen von unmittelbarer Diskriminierung zur Verfügung steht. Kapitel 2 zeigt auch, dass Fragen des Nachweises, der Haftung und der Durchsetzung im Zusammenhang mit der oben beschriebenen Herausforderung der Transparenz und Erklärbarkeit besonders komplex sind. Alles in allem wirft algorithmische Diskriminierung ein neues Licht auf viele der „traditionellen“ Probleme und Kritikpunkte des Gleichstellungs- und Antidiskriminierungsrechts der EU. Wenn die Regulierungsbehörden und politischen Entscheidungsträger der EU wirksamen Rechtsschutz gegen algorithmische Diskriminierung gewährleisten sollen, müssen, so die Argumentation des Berichts, diese Fragen ernst genommen werden – angesichts der oben beschriebenen „Herausforderung des Umfangs und der Geschwindigkeit“ jetzt mehr denn je.

Rechtliche Herausforderungen durch algorithmische Diskriminierung in den europäischen Ländern

Ausgehend von der Beobachtung, dass sich die meisten Beispiele algorithmischer Diskriminierung und ein Großteil der wissenschaftlichen Literatur zu diesem Thema auf den US-amerikanischen Kontext beziehen, widmet sich das dritte Kapitel des Berichts den spezifischen Herausforderungen, die sich durch algorithmische Diskriminierung auf nationaler Ebene in Europa ergeben. Kapitel 3 beginnt mit einer eingehenden Untersuchung des Einsatzes von Algorithmen im öffentlichen und privaten Sektor in 31 Ländern (EU-27, EWR-Länder und Vereinigtes Königreich). In Europa werden Algorithmen von staatlichen Behörden in verschiedenen Bereichen eingesetzt, z. B. in der Arbeitsmarktpolitik (Erstellung von Profilen von Arbeitssuchenden, Zuweisung von Mitteln usw.), der Sozialfürsorge (Vorhersage von Risiken sozialer Ausgrenzung usw.), im Bildungswesen (Einstufung, Auswahl und Zuordnung von Bewerbern zu Hochschulen und Universitäten), in Polizeiarbeit und Betrugsbekämpfung (Aufdeckung von Steuerhinterziehung und -betrug, Prognosen über Kriminalitätsrisiken usw.), in der Justizverwaltung (Unterstützung bzw. Vorausbestimmung der richterlichen Entscheidungsfindung usw.) und in der Regulierung von Medienplattformen (Erkennung und Bekämpfung von Hassreden im Internet usw.). In der europäischen Privatwirtschaft werden Algorithmen u. a. in folgenden Bereichen eingesetzt: Beschäftigung und Plattformarbeit (z. B. in Einstellungsverfahren oder bei der Zuteilung von Arbeit auf Online-Plattformen), Bank- und Versicherungswesen (z. B. zur Prognose von Kreditrisiken und zur Berechnung von Versicherungsschäden) sowie Zielgruppenansprache, Preisgestaltung und Einzelhandel (z. B. zur Personalisierung von Preisen und Angeboten sowie zur Verbreitung von Werbung).

Kapitel 3 geht auch auf die spezifischen Probleme ein, die in diesen Ländern im Zusammenhang mit algorithmenbasierten Entscheidungsprozessen entstehen, und liefert konkrete Beispiele. Die nationalen Expertinnen und Experten berichten in Verbindung mit dem Einsatz von Algorithmen in ihren jeweiligen Ländern über sechs diskriminierungsrelevante Problemkomplexe, die weitgehend den in Kapitel 1 erörterten Merkmalen und Herausforderungen von Algorithmen entsprechen:

- 1) Bias in Daten (z. B. die Verwendung geschützter Merkmale und strukturell verzerrter Daten in Algorithmen, die eingesetzt werden, um Risiken von Arbeitslosigkeit vorherzusagen);
- 2) die diskriminierenden Auswirkungen von Algorithmen (z. B. die stigmatisierenden Effekte einer exzessiven Überwachung von „Problemvierteln“ durch Algorithmen zur Betrugserkennung oder diskriminierendes Verhaltens-Targeting und Personal Pricing);
- 3) Transparenzprobleme und mangelnde Informationen (z. B. die Unmöglichkeit für Gerichte, zivilgesellschaftliche Organisationen, Verbraucher oder staatliche Behörden, Informationen über den möglicherweise diskriminierenden Charakter eines Algorithmus zu erhalten);

- 4) Schwierigkeiten, algorithmische Diskriminierung aufzuspüren und zu identifizieren (insbesondere die Schwierigkeiten für Nutzer und potenzielle Opfer, algorithmische Diskriminierung zu erkennen und – wenn auch nur *prima facie* – nachzuweisen);
- 5) Haftungsfragen (wie zum Beispiel bestimmen, welche der vielen Personen und Einrichtungen, die an der Entwicklung und Anwendung von Algorithmen beteiligt sind, für Diskriminierung verantwortlich ist; außerdem stellen sich Fragen der Zuständigkeit, wenn diese Organisationen auf globaler Ebene etabliert sind);
- 6) die digitale Kluft zwischen den Geschlechtern in Europa (d.h. die eklatante Unterrepräsentanz von Frauen und Minderheitengruppen in Ausbildungsgängen und Berufen in den Bereichen Wissenschaft, Technologie und Ingenieurwesen).

Kapitel 3 geht dann der Frage nach, wie die öffentliche Meinung, die Wissenschaft und die politischen Entscheidungsträger in den 31 Ländern mit diesen Fragen auf nationaler Ebene umgehen. Alles in allem scheint es, dass das öffentliche Bewusstsein für algorithmische Diskriminierung als ein spezifisches Problem, wie auch staatliche Gegenmaßnahmen, in Europa nach wie vor beschränkt sind. In Diskussionen wird algorithmische Diskriminierung oft nur als eines von vielen Problemen gesehen, die generell in Bezug auf Datenschutz und Grundrechte im Bereich der KI bestehen. Auch die nationalen Wissenschaftler neigen dazu, das Problem der Diskriminierung eher unter dem Gesichtspunkt des Datenschutzes als unter dem des Gleichbehandlungsrechts zu betrachten, und juristische Fachliteratur, die sich mit den Herausforderungen befasst, vor die Algorithmen das nationale Antidiskriminierungsrecht stellen, gibt es bislang nur wenig. In einigen europäischen Ländern laufen zwar politische Diskussionen, Berichten der nationalen Expertinnen und Experten zufolge wurden bisher jedoch keine neuen Gesetze oder Gesetzesreformen verabschiedet, um dem Problem der algorithmischen Diskriminierung zu begegnen. In einer Bewertung des in ihren jeweiligen Ländern geltenden Rechtsrahmens stellen die Expertinnen und Experten fest, dass Bestimmungen des Gleichstellungs- und Antidiskriminierungsrechts, des Datenschutzrechts sowie technologie- und branchenbezogene Rechtsvorschriften und Vorschriften des allgemeinen Straf- und Zivilrechts bei der Bekämpfung algorithmischer Diskriminierung zwar eine Rolle spielen können, dass jedoch gewisse Lücken bestehen, die auf nationaler Ebene zu einem problematischen Mangel an Rechtsmitteln führen können. Die festgestellten Lücken sind mit den oben erwähnten identisch und beziehen sich im Wesentlichen auf die Beschränkungen des Geltungsbereichs und auf die Unfähigkeit der bestehenden Rechtsvorschriften, mit den Besonderheiten der algorithmischen Diskriminierung umzugehen. In den meisten europäischen Ländern waren die nationalen Gerichte bislang noch nicht mit algorithmenbasierter geschlechtsbezogener Ungleichbehandlung oder Diskriminierung konfrontiert. Wo Gerichte sich mit Fällen befasst haben, in denen Algorithmen eine Rolle spielten, betrafen diese in der Regel andere Fragen wie z.B. Transparenz und Datenschutz. Wenn auch die Zahl der Rechtsstreite begrenzt ist: Zum Zeitpunkt dieses Berichts sind vor einigen nationalen Gerichten Verfahren anhängig, in denen es um algorithmische Diskriminierung geht, und diese Entwicklungen müssen genau verfolgt werden.

Durchsetzung algorithmischer Gleichheit: Lösungsansätze und Chancen für Geschlechtergleichstellung und Nichtdiskriminierung

Wenngleich der Bericht unterstreicht, dass der Einsatz von Algorithmen zahlreiche Herausforderungen, Risiken und Probleme mit sich bringt, wenn es darum geht, Gleichbehandlung zu gewährleisten und Diskriminierung zu bekämpfen, endet das letzte Kapitel mit einem eher positiven Ausblick. Kapitel 4 zeigt zunächst eine Reihe von Vorteilen und Chancen für Gleichbehandlung und Nichtdiskriminierung auf, die durch den zunehmenden Einsatz von Algorithmen ermöglicht werden. Insbesondere – und im Gegensatz zum menschlichen Gehirn – bieten Algorithmen die Möglichkeit, diskriminierende Verzerrungseffekte besser zu visualisieren, zu messen, zu erkennen und letztlich zu korrigieren, sofern eine angemessene rechtliche Regulierung und Ordnungspolitik vorhanden ist. Algorithmen bieten zum Beispiel neue Möglichkeiten, diskriminierende Stellenanzeigen in großem Maßstab zu erkennen und die Gleichbehandlung von Männern und Frauen in Einstellungsverfahren zu verbessern.

Im Gegensatz zu Kapitel 3, in dem das Fehlen konkreter Gesetzesreformen behandelt wurde, zeigt Kapitel 4, dass es auf nationaler Ebene eine Vielzahl bewährter Verfahren, sowohl öffentliche als auch private, gibt, die darauf abzielen, algorithmische Diskriminierung zu überwachen und zu bekämpfen sowie die betroffenen Berufsgruppen zu diversifizieren. Dazu gehören die Schaffung spezieller Monitoring- und Überwachungseinrichtungen, die Implementierung von Soft-Law-Instrumenten (Ethikkodexe usw.), Selbstregulierungsverfahren (freiwillige Verhaltenskodexe usw.), die Veröffentlichung von Empfehlungen und Leitlinien, die Zusammenarbeit zwischen Datenschutzbehörden und Gleichbehandlungsstellen sowie die Einrichtung öffentlich-privater Allianzen. In Europa wurden auch diverse öffentliche Maßnahmen und private Initiativen zur Diversifizierung von Ausbildungsgängen und Berufen im IT-Bereich ins Leben gerufen – ebenfalls eine wichtige Komponente zur Schaffung diskriminierungsfreier KI. Fast alle diese Maßnahmen zielen jedoch lediglich darauf ab, die Geschlechtergleichstellung zu fördern, anstatt eine breitere Diversifizierung IT-bezogener Ausbildungsgänge und Berufe anzustreben. Tatsächlich würde eine bessere Vertretung aller Minderheitengruppen in diesen Bereichen eine Vielfalt von Sichtweisen begünstigen, was ein entscheidender erster Schritt auf dem Weg zu algorithmischer Gleichheit ist.

Abschließend wird in dem Bericht ein innovativer, integrierter Rahmen vorgeschlagen, der eine Reihe von rechtlichen, wissenschaftlichen und technologischen Maßnahmen und Lösungsansätzen enthält, um algorithmische Diskriminierung zu verhindern, zu bekämpfen und zu beseitigen.

Auf rechtlicher Ebene umfassen diese Maßnahmen vor allem die Verabschiedung des Entwurfs einer horizontalen Richtlinie, über die seit 2008 im Rat verhandelt wird und die den Geltungsbereich des EU-Antidiskriminierungsrechts vereinheitlichen soll, die Schließung der Lücken im Zusammenhang mit den Ausnahmen vom sachlichen Geltungsbereich der Güter- und Dienstleistungsrichtlinie in Bezug auf Medien, Werbung und Bildung sowie eine Klarstellung des Verbots der intersektionellen Diskriminierung. Was die Rolle des Gerichtshofs betrifft, so würde eine weite Auslegung des persönlichen Geltungsbereichs des EU-Gleichbehandlungsrechts – sowohl hinsichtlich des Geltungsbereichs einzelner geschützter Gründe als auch hinsichtlich des abschließenden Charakters der in Artikel 19 AEUV enthaltenen Liste im Lichte der offen formulierten Bestimmung in Artikel 21 der Charta – die Fähigkeit dieses Rechts stärken, gegen algorithmische Diskriminierung vorzugehen. Darüber hinaus wird vorgeschlagen, das Konzept der „Anweisung zur Diskriminierung“ als doktrinelles Ergänzung zu unmittelbarer und mittelbarer Diskriminierung zu verwenden, um eine bessere konzeptionelle Harmonisierung mit algorithmischer Diskriminierung zu gewährleisten und gleichzeitig Verfahrens- und Durchsetzungsprobleme zu verringern.⁹ Was institutionelle Lösungen betrifft, so würde ein öffentlicher, kollektiver Durchsetzungsmechanismus, z.B. eine Europäische Gleichbehandlungsstelle, nach Ansicht des Berichts die Überwachung und Ahndung von algorithmischer Diskriminierung erheblich erleichtern, da ein gerichtliches Vorgehen gegen algorithmische Diskriminierung angesichts der Herausforderung der Transparenz und Erklärbarkeit für einzelne Betroffene derzeit eine unverhältnismäßig hohe Belastung darstellt. Schließlich schlägt der Bericht vor, dass die EU die Schaffung eines Akkreditierungssystems für Zertifizierung und Überwachung in Bezug auf algorithmische Diskriminierung fördern könnte. Ein solches Ökosystem für Zertifizierung könnte einen „Equality-by-Design“-Ansatz fördern, der von Beginn an bei der Konstruktion von Algorithmen zur Anwendung kommt, den Einsatz diskriminierungsfreier Algorithmen seitens öffentlicher und privater Stellen vorantreiben und durch Verbesserung der Transparenz und Erklärbarkeit die Rechtsdurchsetzung erleichtert.

Was wissenschaftliche Lösungen betrifft, so ist es von zentraler Bedeutung, Regulierungsbehörden, Gerichte, Wirtschaftsakteure, den IT-Sektor und die Gesellschaft im Allgemeinen für die Gefahren algorithmischer Diskriminierung zu sensibilisieren. Die Förderung der Forschung zu algorithmischer Diskriminierung wird im Hinblick auf sachlich fundierte Verfahren zur Entwicklung, Anwendung und Regulierung von Algorithmen ebenfalls eine große Rolle spielen. Die Überwachung algorithmischer Diskriminierung sollte durch Meldetools unterstützt werden, die Watchdog-Organisationen und Whistleblower ermutigen könnten, Verdachtsfälle von algorithmischer Diskriminierung zu melden, was wiederum die öffentliche

9 Siehe dazu Xenidis, R., „Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law“ (liegt der Autorin vor).

Aufmerksamkeit auf problematische Praktiken lenken würde. Unerlässlich für die Verhütung und Bekämpfung algorithmischer Diskriminierung sind schließlich auch Ausbildung und Schulung: Analog zur ethischen Ausbildung von medizinischem Personal sollten nicht nur IT-Fachleute, sondern auch relevante Berufsgruppen wie z.B. Regulierungsbehörden, Gerichte oder Gleichbehandlungsstellen hinsichtlich der Gefahren algorithmischer Diskriminierung und der Möglichkeiten, sie zu bekämpfen, ausgebildet und geschult werden.

Zu den technologischen Lösungen schließlich gehört die Umsetzung präventiver Strategien (Folgenabschätzung im Hinblick auf Gleichbehandlung und „Equality-by-Design“-Strategien) in den verschiedenen Phasen der Erstellung von Algorithmen (Entwurf, Training und Entwicklung). Informatiker haben z.B. verschiedene technologische Debiasing-Strategien entwickelt, um algorithmische Diskriminierung sowohl auf der Ebene der Datenauswahl, -kennzeichnung und -verwendung als auch auf der Ebene der algorithmischen Modelle selbst zu minimieren. Technologische Lösungen können auch *a posteriori* zum Zuge kommen, z.B. durch den Einsatz von Screening- und Prüfalgorithmen, die Diskriminierung erkennen können. Derartige technologische Instrumente könnten sich sowohl für die Überwachung algorithmischer Diskriminierung als für die Zertifizierung von diskriminierungsfreien Algorithmen als äußerst nützlich erweisen. Die Entwicklung solcher technologischen Instrumente sollte von den europäischen Behörden unterstützt werden.

Diese verschiedenen rechtlichen, wissensbasierten und technologischen Lösungsansätze sollten bereichsübergreifend integriert werden. Der Bericht schlägt dafür den Rahmenplan **PROTECT** vor, eine Reihe von Schlüsselempfehlungen für staatliche Maßnahmen in Europa, die um sieben Handlungsschwerpunkte herum strukturiert sind:

- ▶ **PREVENT (VERHINDERN)**: Mithilfe von vielfältigen, gut ausgebildeten IT-Teams, Folgenabschätzungen im Hinblick auf Gleichbehandlung, Ex-ante-„Equality-by-Design“- bzw. Ex-ante-„Legality-by-Design“-Strategien.
- ▶ **REDRESS (ABHELFFEN)**: Kombination verschiedener Instrumente des Antidiskriminierungsrechts, des Datenschutzrechts usw., um eine eindeutige Zuordnung rechtlicher Verantwortlichkeiten, klare Rechtsbehelfe, faire Beweisregeln sowie eine flexible und reaktionsfähige Auslegung und Anwendung von Nichtdiskriminierungskonzepten zu fördern.
- ▶ **OPEN (ÖFFNEN)**: Förderung von Transparenz, etwa durch Vorgaben zur Verfügbarmachung offener Daten zu Kontrollzwecken (z.B. Zugang zu Quellcodes).
- ▶ **TRAIN (SCHULEN)**: Ausbildung, Schaffung und Verbreitung von Wissen über Nichtdiskriminierungs- und Gleichbehandlungsfragen unter IT-Fachleuten, Sensibilisierung von Regulierungsbehörden, Richtern, Personalverantwortlichen, Staatsbediensteten und der Gesellschaft im Allgemeinen für Fragen der algorithmischen Diskriminierung.
- ▶ **EXPLAIN (ERKLÄREN)**: Festlegung von Anforderungen bzgl. Erklärbarkeit, Rechenschaftspflicht und Informationen.
- ▶ **CONTROL (KONTROLLIEREN)**: Aktive Beteiligung des Menschen (menschenzentrierte KI), etwa in Form von HITL-Systemen (Human in the Loop), die ein automatisches Genehmigen verhindern, ergänzt durch Supervisions- und Konsultationsmechanismen (Kontroll- und Konsultationskette mit Nutzern).
- ▶ **TEST (PRÜFEN)**: kontinuierliche Überwachung von Hoch-Risiko-Algorithmen und deren Output, Einrichtung von Prüf-, Kennzeichnungs- und Zertifizierungsmechanismen.

General introduction

Subject, context and scope of the report

The rapid development and increasing use of artificial intelligence (AI) and algorithmic applications has raised many concerns relating to the propensity of algorithms to discriminate. Algorithmic discrimination – a phenomenon also often called ‘algorithmic bias’¹ – can arise from various sources and endanger one of the most fundamental rights guaranteed by EU law: the right to gender equality and non-discrimination. This has recently been recognised by the European Commission in its white paper on artificial intelligence, which emphasises that ‘[a]rtificial intelligence (AI) entails a number of potential risks, such as [...] gender-based or other kinds of discrimination’.² Underlining the specific risks to gender equality, the recent European Commission Gender Equality Strategy 2020-2025 warns that ‘[w]hile AI can bring solutions to many societal challenges, it risks intensifying gender inequalities’ and that ‘[a]lgorithms and related machine-learning, if not transparent and robust enough, risk repeating, amplifying or contributing to gender biases that programmers may not be aware of or that are the result of specific data selection’.³

The risks of algorithmic discrimination are pervasive and occur at various stages, from programming, building, training and testing to operating algorithms. For example, machine-learning (including deep-learning) algorithms, which rely on correlations identified in large amounts of social and personal data, may reproduce patterns of past inequalities.⁴ Discrimination may also occur in rule-based algorithms.⁵ For instance, human biases, prejudices and stereotypes may influence the design, parameters and rules of algorithms at the phase of developing an algorithm, and the outcome might be discriminatory. This is highly problematic from the perspective of equality and non-discrimination law, in particular when discriminatory algorithms are used for automated decision making, or in support of human decision making. What is more, the lack of awareness about existing risks of discrimination adds to the presumption of neutrality associated with technology, potentially leading to the invisibility of algorithmic discrimination.⁶

These issues raise the question of the adequacy of the legal framework in place to protect citizens from algorithmic discrimination. Attention has recently been drawn to potential inadequacies in the existing anti-discrimination legal framework at EU level.⁷ For example, EU law covering the majority of protected grounds is limited to the employment sphere, while many algorithmic applications are deployed in other areas of life such as retail, healthcare, insurance and the like, as well as for Government purposes (such as predictive policing, distribution of social benefits and taxation). In addition, questions arise regarding the interaction, adequacy and complementarity of EU non-discrimination law and EU legislation on the protection of personal data and their ability to effectively protect citizens against discrimination, in

-
- 1 See e.g. Hacker, P (2018) ‘Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law’ 55 *Common Market Law Review* 1143; Angwin J and others (2016), ‘Machine Bias’, *ProPublica*, 23 May 2016, available at: www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Mayson, SG (2018), ‘Bias In, Bias Out’, University of Georgia School of Law Legal Studies Research Paper No 2018-35. In more detail, see also section 1.5. of this report.
 - 2 European Commission (2020), ‘White Paper on Artificial Intelligence’.
 - 3 European Commission (2020), Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions ‘A Union of Equality: Gender Equality Strategy 2020-2025’ COM(2020) 152 final (Brussels 2020).
 - 4 See e.g. Barocas, S and Selbst, AD (2016), ‘Big Data’s Disparate Impact’ 104 *California Law Review* 671 and Buolamwini, J and Gebru, T (2018) *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification* (Proceedings of Machine Learning Research). For more information about machine-learning algorithms, see section 1.2.2.
 - 5 On rule-based algorithms, see further section 1.2.1.
 - 6 This phenomenon is called automation bias, see e.g. Skitka, LJ, Mosier, KL and Burdick, M (1999) ‘Does automation bias decision-making?’ 51 *International Journal of Human-Computer Studies*. In more detail, see section 1.4.1 of this report.
 - 7 See e.g. Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’ 55 *Common Market Law Review* 1143 and Xenidis, R and Senden, L (2020), ‘EU Non-discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination’, in Bernitz, U and others (eds), *General Principles of EU Law and the EU Digital Order* (Alphen aan den Rijn, Kluwer Law International).

particular in relation to the collection and processing of sensitive personal data as regulated under the General Data Protection Regulation.⁸

This report therefore investigates how algorithmic discrimination challenges the set of legal guarantees put in place in Europe to combat discrimination and ensure equal treatment. More specifically, it examines whether and how the current EU gender equality and non-discrimination legislative framework can adequately capture and redress algorithmic discrimination. It explores the gaps and weaknesses that emerge at both the EU and national levels from the interaction between, on the one hand, the changing and specific types of discrimination that arise when algorithms are used in decision-making systems and, on the other, the particular material and personal scope of the existing legislative framework and the concepts of discrimination that it relies on. The report also maps out existing legal solutions, accompanying policy measures and good practice to address and redress algorithmic discrimination both at EU and national levels. Finally, the report proposes its own integrated set of legal, knowledge-based and technological solutions to the problem of algorithmic discrimination.

Problems of bias and discrimination in and by algorithms and algorithm-based decision making arise in relation to all grounds of discrimination. While this report adopts a specific focus on gender equality, it is scientifically grounded in general non-discrimination legal theory. For that reason, it focuses on (1) general questions that are relevant to all protected grounds of discrimination under EU law, including gender-based discrimination and (2) specific gender equality issues and questions. There are two main motivations for this twofold approach. First, a large number of questions pertaining to algorithmic discrimination are common to all grounds protected under EU discrimination law and therefore it makes sense to approach the topic of this report from a general equality law perspective. However, specific issues arise in relation to gender equality and gender-based discrimination, which is why challenges, risks and questions related to gender bias are a special focus in this report where relevant. Secondly, this approach is in line with the implementation of the Gender Equality Strategy 2020-2025, which uses 'intersectionality – the combination of gender with other personal characteristics or identities, and how these intersections contribute to unique experiences of discrimination – as a cross-cutting principle'.⁹ As acknowledged by the strategy, 'the intersectionality of gender with other grounds of discrimination [should] be addressed across EU policies [because] [w]omen are a heterogeneous group and may face intersectional discrimination based on several personal characteristics'.¹⁰ The prevalence of intersectional discrimination might even increase with algorithm-based decision making in light of data-driven profiling techniques, such that gender inequality should be considered in relation to other protected grounds. For example, in many cases the output of an algorithm and the decision based on it will not be based *only* on sex or *only* on ethnic origin, but on a combination of characteristics and behaviours that is unique to a particular person or a small group of persons, without it being particularly clear which of these characteristics was most important in making that decision. As a result, a study on algorithmic discrimination necessarily has to combine a general non-discrimination approach with specific attention to gender equality issues.

Because of the scarcity of dedicated legal instruments and case law,¹¹ this report adopts a slightly different approach than other EELN thematic reports that have been written on areas of EU gender equality and non-discrimination law that are already well established. One of the main goals of this report is to frame the problem of algorithmic discrimination in the particular context of EU law and set the stage by delineating major problems, gaps and weaknesses in the current legal instruments at

8 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1.

9 European Commission (2020), Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'A Union of Equality: Gender Equality Strategy 2020-2025' COM(2020) 152 final (Brussels 2020), p. 2.

10 European Commission (2020) 'A Union of Equality: Gender Equality Strategy 2020-2025', p. 16.

11 This scarcity is further discussed in section 3.4.

EU and national level. This is a necessary step given the current lack of an overview.¹² In addition, this report aims to identify and analyse relevant legal debates and discussions, existing legal instruments and policy initiatives as well as case law, when available, both in the European states and at EU level. This account provides a useful overview of existing legal challenges in relation to algorithmic discrimination, the various ways they are perceived, understood, framed and approached, and the solutions that are being discussed or implemented in academia and practice.

Methodology

The methodology underlying this report is twofold. First, this report offers a general theoretical discussion and a legal analysis of the characteristics and challenges of algorithm-driven decision making, as well as the risks and problems of algorithmic discrimination. This part of the analysis focuses on the technological, legal, conceptual, doctrinal and enforcement challenges that the use of algorithms create in the context of gender equality and non-discrimination law. It relies on legal and interdisciplinary scholarly literature in which relevant problems, challenges and solutions are identified and discussed. Where relevant and possible, it is complemented by a legal and case law analysis. Because of the limited availability of related case law and legislation at this early stage, the analysis often proceeds by analogy with existing law and doctrine. By critically analysing case law and legal provisions in light of the new angle of algorithmic discrimination, original and critical insights are offered into the strengths and weaknesses of the current EU legal framework and the equality *acquis*.

Secondly, the report maps out and analyses how various actors in Europe have responded to problems of algorithmic discrimination. To do so, a review has been made of policy and regulatory sources of European relevance in matters of AI and gender equality and non-discrimination. These sources include official EU and Council of Europe documents, such as the European Commission's Communications on AI,¹³ the EU white paper on AI,¹⁴ the Gender Equality Strategy 2020-2025,¹⁵ and the Ethics Guidelines for Trustworthy Artificial Intelligence of the High-Level Expert Group on AI.¹⁶ This report also maps out relevant discussions, challenges and policy-making efforts at the national level in 31 countries in Europe: the EU-27, the three EEA countries and the United Kingdom. Much of this information was gathered from a detailed questionnaire, containing 11 sets of thematic questions, which is reproduced in the annex of this report. The questionnaire was sent to the national experts in gender equality law of the European Network of Legal Experts in gender equality and non-discrimination law, who collaborated with the national experts in non-discrimination law in order to respond to the questionnaire and offer their analysis of the national law, case law and policy and scholarly discussions. The thematic and comparative analyses offered in Chapters 3 and 4 are mostly based on their answers and insights.

Structure

Chapter 1 provides the background to the report and sets the scene for the discussion of the various problems raised by algorithms in relation to gender equality and non-discrimination law. It offers key definitions, terminological clarifications as well as an introduction to the use that can be made of the different types of algorithms. It introduces readers to different types of algorithmic technologies and their characteristics, as well as to the various stages at which bias can impact algorithms from design to use and from planning to development. This introduction is mainly addressed to readers who are not familiar

12 So far, the problem of algorithmic discrimination has mainly been addressed from an ethical perspective and insufficiently from a legal one, and when this was the case, legal studies have overwhelmingly focused on the United States.

13 European Commission (2018), Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on 'Artificial Intelligence for Europe' COM/2018/237 final (Brussels 2018); Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 'Building Trust in Human Centric Artificial Intelligence' COM(2019)168 (Brussels 2019).

14 European Commission (2020) 'White Paper on Artificial Intelligence'.

15 European Commission (2020), 'A Union of Equality: Gender Equality Strategy 2020-2025'.

16 High-Level Expert Group on AI (2019), 'Ethics Guidelines for Trustworthy Artificial Intelligence' (European Union: Brussels 2019).

with the different types of algorithmic technologies and their characteristics – readers who are familiar with them might skip section 1.2 and jump directly to the next section, which takes the reader through the various phases in which discrimination can creep into algorithms. The first chapter also examines what is new and different about *algorithmic* discrimination as opposed to *human* discrimination. Importantly, this chapter delineates six major challenges posed by algorithms in the context of gender equality and non-discrimination law: (1) the human factor and the stereotyping and cognitive bias challenge, (2) the data challenge, (3) the correlation and proxies challenge, (4) the transparency and explainability challenge, (5) the scale and speed challenge and (6) the responsibility, liability and accountability challenge. Finally, the chapter aims to offer conceptual clarification by explaining the difference between notions of algorithmic bias, fairness and discrimination, and it explains the interrelationship between algorithmic discrimination and the protection of personal data.

The second chapter of the report maps out how the types of discrimination arising from the increasing use of algorithms represent challenges for EU gender equality and non-discrimination law. It reviews the whole EU legal framework protecting equality in light of the specific risks of discrimination that the use of algorithms poses in various contexts. The gaps, shortcomings and weaknesses of the core elements of the EU non-discrimination legal regime are closely examined and analysed in light of the six challenges outlined in Chapter 1. The chapter highlights how algorithmic discrimination risks falling into the cracks of EU gender equality and non-discrimination law because of current gaps in the material scope, uncertainties and lack of flexibility in the personal scope, conceptual frictions, doctrinal mismatches, procedural difficulties and enforcement challenges.

Chapter 3 analyses the legal challenges that algorithmic discrimination poses to national equality law. Based on the national experts' reports, it maps out how algorithms are used in the public and private sectors in the 31 European countries studied in this report, offering specific examples and illustrations. The aim is to offer a counterpoint to US-centric discussions of problems of algorithmic discrimination by outlining the specific challenges that arise in the European context. This chapter highlights six specific sets of discrimination issues that experts in the 31 countries perceive to exist in relation to the use of algorithms, which resonate both with the EU-wide challenges explained in this report and the general scientific literature on algorithmic discrimination: (1) biases in data, (2) the discriminatory effects of algorithms, (3) transparency problems and lack of information, (4) difficulties in detecting and identifying algorithmic discrimination, (5) responsibility issues and (6) the gender digital gap in Europe. Chapter 3 also reviews how and to what extent the problem of algorithmic discrimination is framed in national public discussions, legal scholarship and policy debates. Based on national experts' evaluation of national gender equality and non-discrimination law, data protection law as well as technology-specific legislation, sectoral legislation and general criminal and civil law provisions, Chapter 3 identifies the specific gaps and shortcomings that could lead to a lack of remedies at national level.

Finally, the last chapter of the report focuses on benefits, opportunities, good practice and solutions, and proposes an integrated framework of measures to prevent and combat algorithmic discrimination. It discusses how algorithms can offer opportunities to better visualise, measure, detect and ultimately correct discriminatory biases if proper legal regulation and public policy is put in place. It also offers insights into a wide range of public and private good practice adopted at national level to monitor and address algorithmic discrimination – from the creation of monitoring bodies to voluntary codes of conduct and from recommendations, guidelines and ethics codes to cooperation between data protection agencies and equality bodies – and to diversify relevant professional communities with a view to achieving more equality through better representation and participation of minority groups in the design and development of algorithms. The report closes by proposing a new integrated framework called PROTECT, which offers a set of legal, knowledge-based and technological measures and solutions to prevent, address and redress algorithmic discrimination.

1 What is algorithmic discrimination and what is new about it?

1.1 Introduction

Algorithms can be defined as a set of computer instructions that, based on a series of input data, can produce a certain value or set of values as output.¹⁷ Some algorithms can directly inform a decision, such as a decision to grant a social security benefit to a specific person, or to fine someone who has been found to be speeding.¹⁸ Other algorithms mainly calculate probabilities, such as the probability that a certain deviation in human cells is indicative of cancer, a person with certain qualifications is suited for a particular job, or a certain type of crime will be committed in a certain neighbourhood.¹⁹ Such algorithms usually do not directly inform decisions, but they can support decision making by human beings. For example, a doctor may take account of the probability calculations made by the algorithm in helping her diagnose cancer, or the police may decide to patrol a certain neighbourhood more intensively. These examples readily show that there are different types of algorithms, and that they can have different functions.

Since the differences between algorithms may matter to the way in which they can be dealt with by gender equality and non-discrimination law, they are explained in section 1.2. This section also shows that, depending on their characteristics, the different types of algorithms can have various functions, such as automated decision making, pattern detection, profiling, classification, clustering, and probability calculations.²⁰

For the purposes of this study, algorithms are mainly relevant insofar as they generate or support decisions, ranging from product pricing to the medical diagnosis of a certain disease and from making a personalised suggestion to watch a certain television series to the decision to hire someone for a job. To help identify and locate specific problems of discrimination or bias in algorithmic decision making, as well as issues of responsibility and accountability, it is helpful to distinguish different stages in the process of algorithm-based decision making.²¹ These stages are discussed in section 1.3, which also provides various examples of possible uses of algorithms.

Section 1.4 constitutes the core section of the current chapter, as it highlights the main challenges for equal treatment and non-discrimination of algorithms, which follow from their specific and shared characteristics. As is explained in this section, algorithms require significant human intervention, their quality and reliability depend to a large extent on the quality and reliability of the data that are being used to train and feed them, they help detect correlations and patterns rather than causal relations, and they are commonly non-transparent and their operation is difficult to explain to lay people (and sometimes even to technology experts). Other challenges that are closely related to these characteristics and that are relevant to gender equality and non-discrimination relate to the scale and speed of algorithmic decision making, as well as to determining who is responsible for algorithmic discrimination.

17 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek* (Legal aspects of algorithms that make decisions. An exploratory study) (WODC/Utrecht University Utrecht) section 2.1; Cormen, TH and others (2009), *Introduction to Algorithms* (Cambridge, MIT Press) p. 5. See also Gillespie, T (2014), 'The Relevance of Algorithms' in Gillespie, T, Boczkowski, PJ and Foot, KA (eds), *Media technologies: Essays on communication, materiality, and society* (Cambridge, MIT Press) p. 167.

18 See Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.2.

19 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.

20 See also Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten* (Algorithms and fundamental rights) (Den Haag, Boom Juridisch) p. 20.

21 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 8.2.

Finally, for the purposes of clarity, some differences between technological and ethical terminology and legal terminology are highlighted in Section 1.5. In particular, the notions of algorithmic fairness and algorithmic bias are set out in relation to the legal notions of equality and non-discrimination. In addition, the interactions between data protection law and equality and non-discrimination law are briefly explained, as well as the notion of ‘sensitive personal data’ used in the EU’s General Data Protection Regulation.

The chapter closes with a short concluding summary in section 1.6.

1.2 Types of algorithms²²

1.2.1 Rule-based algorithms

Most decision-making procedures require logical thought processes that can be simplified to reasoning that follows the pattern: ‘if this, then that’.²³ To give a simple example, a legislative rule may state that driving faster than 100 km per hour on a motorway is prohibited and that violation of that rule is punishable by a fine of EUR 150. Consequently, *if* someone has been found to have been driving faster than 100 km per hour, *then* the consequence must be that she has to pay a EUR 150 fine. Of course, in most cases, decision-making processes are much more complex than this, especially because they consist of several logical (sub-)rules and include many more variables. Nevertheless, if the rules are sufficiently clear and the variables are well known, even such complex thought processes can be split into different ‘if this, then that’ constructions, resulting in (often highly complicated) decision trees.

Such ‘if this, then that’ processes and the resulting decision trees can be relatively easily translated into computer instructions or algorithms.²⁴ This results in *rule-based algorithms* or ‘knowledge-based’ systems that are highly predictable, since the set of instructions and rules is fixed and all possible variables and outcomes are programmed into the algorithm.²⁵ Once fully developed, such rule-based algorithms can therefore relatively easily replace human decision making: humans only need to feed the algorithm with relevant data, and the algorithm can then automatically produce output in the shape of a decision that is in line with the decision-making process that humans have devised.

The main *function* of such rule-based algorithms is to speedily handle large volumes of decisions, ensure consistence and reduce the number of human mistakes in conducting repetitive tasks.²⁶ However, such algorithms cannot always be used – to program a rule-based algorithm, precise, clear-cut and logical rules are needed with clear sets of unambiguous variables that can be translated into non-ambivalent ‘if this, then that’ rules and sub-rules. Usually, substance experts (such as lawyers) work together with technical experts to identify the different aspects of a decision-making process to see whether and how it can be translated into digital choices (‘either this or that’) and how the logic of the process can be programmed into a computer instruction.²⁷

22 This section makes use of the definitions and distinctions developed in a report to which one of the present authors contributed – Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek* (Legal aspects of algorithms that make decisions. An exploratory study) (WODC/Utrecht University).

23 Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten* (Algorithms and fundamental rights) (Den Haag, Boom Juridisch) 48.

24 See further Surden, H (2019), ‘Artificial Intelligence and Law: An Overview’, 35 *Georgia State University Law Review* 1306, 1316, discussing logical rules and knowledge representation.

25 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.1; Surden, H (2019), ‘Artificial Intelligence and Law: An Overview’, 35 *Georgia State University Law Review* 1306, 1316; Raso, FA and others (2018), ‘Artificial Intelligence & Human Rights: Opportunities & Risks’, The Berkman Klein Center for Internet & Society Research Publication Series, available at: <https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights>, 10.

26 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.1; Surden, H (2019), ‘Artificial Intelligence and Law: An Overview’, 35 *Georgia State University Law Review* 1306, 1318.

27 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.1; Surden, H (2019), ‘Artificial Intelligence and Law: An Overview’, 35 *Georgia State University Law Review* 1306, 1316.

1.2.2 Machine-learning algorithms

Different from rule-based algorithms, machine-learning algorithms are characterised by their ability to 'learn', that is to autonomously adapt, evolve and improve in order to optimise any given outcome based on any input data without being explicitly programmed to do so. Rule-based algorithms are *static*, because their rules can only be altered through programming, but machine-learning algorithms can be considered *dynamic* because their rules change depending on the input data.

1.2.2.1 Data mining: techniques for analysing (big) data

Algorithms are used not only to program instructions for decision making and decision trees into an automatised system, but can also be used for pattern recognition, profiling or probability calculations. In particular, machine-learning algorithms can be applied in this way.²⁸ Such machine-learning algorithms make use of different analytical instruments and techniques, which all focus on finding correlations and patterns in (big) data, as outlined below.²⁹

- 1) *Classification techniques* – based on pre-defined categories or 'classes', an algorithm can be trained to detect which data belong to which categories (see below, under *supervised learning*).³⁰ Such classification techniques are used, for example, in spam detection.³¹ An algorithm can be trained to recognise spam on the basis of what humans generally know about typical 'spam' messages and their wording. Experts can identify different classes or categories of messages that can generally be recognised as spam. They can then train an algorithm to recognise such categories in large volumes of incoming emails and teach it to search for similar patterns and wording that it has learned to recognise as spam.³² Once fully functional, the algorithm can then independently detect typical spam patterns, and autonomously decide to place the recognised emails in a spam folder.³³
- 2) *Clustering techniques* – an algorithm can learn to identify strong commonalities or correlations between seemingly highly diverse data. This allows it to create clusters of situations, or, for example, of persons with comparable interests, preferences or capacities.³⁴ Clustering techniques can be used to detect commonalities, but also to identify outliers – situations that are oddly unlike other types of cases. Such clustering techniques can be used, for example, in detecting fraudulent tax reports, which may reveal themselves by having atypical characteristics that make them stand out from regular reports.³⁵
- 3) *Regression techniques* – regression techniques can be used to calculate probabilities. For example, banks may calculate credit risks by comparing personal data (such as someone's credit history and personal situation) to all available data to be able to estimate the likelihood that that person will

28 On machine learning, see e.g. Hamon, R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* (JRC Technical Report, EU Science Hub) (European Commission, Brussels) 10; Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 35 *Georgia State University Law Review* 1306, 1311. In this report we distinguish between machine-learning and deep-learning algorithms for reasons of clarity, but these are sometimes also regarded as part of the same category. For the same categories, see also e.g. Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making: Opportunities and challenges* (Panel for the Future of Science and Technology (STOA) of the European Parliament), 4.

29 Barocas, S and Selbst, A (2016), 'Big Data's Disparate Impact' 104 *California Law Review* 761, 677; Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten* (Algorithms and fundamental rights) (Den Haag, Boom Juridisch) 20-21. For a definition of 'big data', see e.g. Information Commissioner's Office (ICO) (2017), *Big data, artificial intelligence, machine learning and data protection*, available at: <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.

30 Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 35 *Georgia State University Law Review* 1306, 1312; Barocas, S and Selbst, A (2016), 'Big Data's Disparate Impact' 104 *California Law Review* 761, 678.

31 Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 1306, 1312.

32 Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 1306, 1312.

33 Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 1306, 1314.

34 Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten*, 22.

35 Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten*, 22.

repay a loan.³⁶ Such regression techniques can also be built into an algorithm, which then can be trained to search big data for probabilities and regressions with great speed.³⁷

- 4) *Association techniques* – an algorithm can be trained to search for particular correlations between data that suggest future behaviour. For example, the algorithm may detect that buying a smartphone is closely correlated to buying a smartphone cover, or that most people who have watched a particular series then continue to watch another series. These correlations can be translated to ‘association rules’ that can, for example, be used to make suggestions to customers or individual users: if the user is interested in A, he may also be interested in B.³⁸ In addition, the predictions that such association rules and correlations make can help in setting prices, in selection procedures or in personalising information and behavioural targeting.

1.2.2.2 Functions

Using combinations of the techniques discussed above, self-learning algorithms can be trained to have different functions. They can be used, first, for *profiling* purposes.³⁹ For example, the clustering and association techniques can help create a profile of a particular person (e.g. a male nurse who is unmarried, cares for a two-year old child, likes running and is keen on horror movies), and then, based on regression techniques and probability calculations, the algorithm can make predictions as to other, yet unknown, preferences that this particular person might have (e.g. a special preference for vegetarian dishes). Similar techniques can be used to create group profiles, e.g. identify groups with common attributes that can be seen to predict a particular risk of developing a certain illness or an inclination to radicalise.

The above example also shows that self-learning algorithms may have *predictive* functions.⁴⁰ Based on pattern recognition, identification of correlations between data and probability calculations and regression analyses, an algorithm may help predict how a certain individual will behave in the future or which events or consequences of a certain act are likely to occur.⁴¹

1.2.2.3 Supervised learning

Algorithms cannot make any of the analyses or have any of the functions discussed above from their own motion. They need to be specifically developed and ‘trained’ to analyse data in a specific way, trace correlations, find relevant patterns, make calculations, predict future behaviour, create individual profiles and the like. There are different available ways of doing so. The first method is ‘supervised learning’, which is often used in relation to classification of data (as in the example of the spam filtering systems discussed above).⁴² Put simply, an algorithm is fed carefully selected and pre-categorised ‘labelled data’ (e.g. messages that clearly contain different forms of spam) and is instructed that these data disclose a

36 Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten*, 22.

37 See e.g. Larus, J and others (2018), ‘When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making’ (Technical Report Informatics Europe & EUACM) 6.

38 Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten*, 22.

39 Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten*, 22; Mendoza, I and Bygrave, LA (2017), ‘The right not to be subject to automated decisions based on profiling’ in Synodinou, T and others (eds), *EU Internet Law* (Berlin/Heidelberg, Springer) 77; Schermer, B (2011), ‘The limits of privacy in automated profiling and data mining’ 27 *Computer Law & Security Review* 45; Hildebrandt, M (2008), ‘Defining profiling: A new type of knowledge?’ in Hildebrandt, M and Gutwirth, S (eds), *Profiling the European citizen* (Springer, Rotterdam) 17.

40 See further eg Larus, J and others (2018), ‘When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making’ (Technical Report Informatics Europe & EUACM), 7; Schermer, B (2011), ‘The limits of privacy in automated profiling and data mining’ 27 *Computer Law & Security Review* 45.

41 Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten*, 22-3; Gillespie, T (2014), ‘The Relevance of Algorithms’ in Gillespie, T, Boczkowski, PJ and Foot, KA (eds), *Media technologies: Essays on communication, materiality, and society* (Cambridge, MIT Press) 167, 173-4.

42 Information Commissioner’s Office (ICO) (2017), *Big data, artificial intelligence, machine learning and data protection*, available at: <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>, para. 10. See also Huq, AZ (2020), ‘A Right to a Human Decision’ 105 *Virginia Law Review* 15.

certain category or class.⁴³ New data are then fed into the system and the algorithm is asked to recognise the same or similar patterns in those data. This is called the ‘training’ phase. If the algorithm recognises the relevant patterns in the labelled data correctly (according to the data scientists’ input), it is given positive feedback, whereas it gets negative feedback if it does not categorise the data correctly. Based on this feedback, the algorithm will gradually correct and improve itself, up to a point that it satisfactorily recognises the relevant categories in the data presented to it. Once that point is reached, the algorithm can be validated. This means that it can then safely operate outside the highly controlled context of a data lab and without the pre-labelled data and can help to assist decision-making processes in the real world. This process of training, giving feedback and learning by the algorithm is called *supervised machine learning*, since data scientists are very closely involved in deciding whether the algorithm detects the correct patterns and give feedback on that basis.

1.2.2.4 Unsupervised learning

Other algorithms are based on unsupervised learning.⁴⁴ These algorithms are provided with a set of instructions and a large amount of training data in which they are asked to discover correlations and patterns autonomously, mainly using clustering and regression techniques.⁴⁵ The difference to supervised learning is that these algorithms are not trained with labelled data.⁴⁶ It is therefore not possible to constantly check whether they discover patterns and correlations in the way that human beings would.⁴⁷ In unsupervised machine learning, the only checks that can be made are related to the algorithm’s output, which can be seen to either comply or not comply with human expectations.⁴⁸ This is the reason why these algorithms are often called ‘black box’ algorithms. On the basis of the output, feedback can be given to the algorithm, which can then use that feedback to correct itself if needed.

1.2.2.5 Reinforcement learning

One last form of machine learning is reinforcement learning.⁴⁹ This is where an algorithm navigates in a certain environment (e.g. a financial trading system or an employment context) and is instructed to achieve a certain aim (e.g. optimising a transaction or finding the best-suited candidate for a certain job).⁵⁰ If the aim is achieved the algorithm will be given positive feedback; if it is not, the feedback will be negative. Eventually, using regression analysis, calculations of probabilities and recognition of correlations and patterns, the algorithm can learn which of a wide range of conceivable scenarios is related most closely to the aim that it is asked to realise, and which actions or decisions would best contribute to achieving that aim.⁵¹

43 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek* (Legal aspects of algorithms that make decisions. An exploratory study) (WODC/Utrecht University Utrecht), section 2.1.2; Hamon, R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* (JRC Technical Report, EU Science Hub) (European Commission, Brussels) 10; Information Commissioner’s Office (ICO) (2017), *Big data, artificial intelligence, machine learning and data protection*, para 10. For the risks of discrimination involved in this process, see Barocas, S and Selbst, A (2016), ‘Big Data’s Disparate Impact’ 104 *California Law Review* 761, 681.

44 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.2.

45 See Huq, AZ (2020), ‘A Right to a Human Decision’ 105 *Virginia Law Review* 16.

46 Hamon, R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* (JRC Technical Report, EU Science Hub) (European Commission, Brussels) 10.

47 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.2. See also Information Commissioner’s Office (ICO) (2017), *Big data, artificial intelligence, machine learning and data protection*, para 10.

48 See e.g. Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’ 55 *Common Market Law Review* 1143, 1147; Kroll, JA, Huey, J and others (2017), ‘Accountable Algorithms’ 165 *University of Pennsylvania Law Review* 633, 645.

49 Hamon, R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence*, 10.

50 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.2.

51 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.2. See also Hamon, R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence*, 10.

1.2.3 Deep learning

The most advanced algorithms from a technological perspective are deep-learning algorithms.⁵² These algorithms learn to operate in roughly the same way as the human brain does, that is, by using so-called neural networks.⁵³ For example, a deep-learning algorithm may learn to divide complex processes of pattern recognition in different layers (a 'multi-layered perceptron').⁵⁴ Layering helps it to split up a complex element into smaller units which can then be more easily identified and processed. For instance, image recognition techniques rely on deep-learning algorithms, which in order to identify the content of a given image (e.g. a dog, a face or a tree), might analyse the different colour elements in the image, the arrangement of pixels in a given area, contrast patterns, etc. By combining these various bits of information, the algorithm might be able to predict the correct outcome, i.e. what the image is. Deep-learning algorithms differ from other types of machine-learning algorithms in that they are able to identify patterns in new data without being extensively trained with selected datasets and without being given human feedback on their output.⁵⁵ They are currently mainly used for complex tasks such as image, speech and facial recognition and automated translation.⁵⁶

1.2.4 Enabling technologies and combining algorithms: AI

In modern societies, all types of algorithms that have been discussed above are being used in different ways and contexts; they serve different functions and rely on different (combinations of) technologies. For example, rule-based algorithms are frequently used in decision-making procedures, machine-learning algorithms may help predict risks of social security fraud or assist in interpreting CT-images after someone has suffered a stroke, and deep-learning algorithms are invaluable to image and facial recognition technology. It is important to note that in order to serve these functions, many current applications rely on connections with other technologies, in particular with *enabling technologies*.⁵⁷ Such enabling technologies can serve to generate the large amounts of data (*big data*) that algorithms need to identify patterns and correlations and calculate probabilities. For example, an increasing number of 'things' (ranging from smartphones to dolls and light switches) are nowadays equipped with sensors, cameras or microphones, allowing them to detect movement, sound or weight. If these 'things' are connected to the internet, they can collect and transfer the information their sensors have detected. The big data generated by this 'internet of things' (IoT) can then be used to train and optimise self-learning algorithms, as well as use them in making predictions or offering suggestions.

In addition, (different types of) algorithms may interact.⁵⁸ For example, one algorithmic process can be instructed to make use of the output generated by another algorithm.⁵⁹ This allows for increasingly complicated, interdependent and mutually influencing processes of data analysis and identification of correlations and patterns.

52 See e.g. Hamon, R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence*, 10. For much more information, see Goodfellow, I, Benjo, Y and Courville, A (2016), *Deep Learning* (Cambridge, MIT University Press).

53 Marr, B (2016) 'What is the difference between deep learning, machine learning and AI?' *Forbes* (8 December), available at: www.forbes.com/sites/bernardmarr/2016/12/08/what-is-the-difference-between-deep-learning-machine-learning-and-ai/#9968fb626cfa.

54 See e.g. Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* 19; see also <https://deepai.org/machine-learning-glossary-and-terms/multilayer-perceptron>.

55 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.2.

56 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.2.

57 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.2.

58 Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 35 *Georgia State University Law Review* 1306, 1319, speaking of 'hybrid' systems.

59 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.2.

Finally, algorithms may be integrated into other physical or cyber systems, such as automated cars, autonomous weapons or domestic or care robots.⁶⁰ These possibilities for combining algorithms and other emerging digital technologies are opening up unprecedented opportunities, yet their interdependency and interaction may also create particular difficulties and risks.⁶¹

To describe such interconnected processes that allow algorithms to perform tasks autonomously in a way that is close to what humans can do, the term artificial intelligence or AI is often used.⁶² As Surden has emphasised, however, AI systems operate by using computational mechanisms that do not resemble or match actual human thinking.⁶³ In addition, most AI tends to be 'narrow' intelligence, which means that it is tailored only to undertake a limited number of very specific tasks.⁶⁴ Moreover, AI mechanisms and algorithms have several distinct characteristics that set them apart from human intelligence and decision making, and which are described further in section 1.4 below. These characteristics pose some specific challenges from the perspective of equality and non-discrimination, as is also addressed in section 1.4.

1.3 Stages of algorithmic decision making and uses of algorithms

Kulk, Van Deursen and others have shown that, generally, three stages can be distinguished in processes of algorithmic decision making, regardless of the type of algorithm that is being used.⁶⁵ These distinctions are useful for the purposes of this study, since different risks of discrimination can be seen to be involved in each of the three stages and different actors can be held responsible for such instances of discrimination.⁶⁶ In other words, discrimination can infect algorithms from their inception to their end use, leading to problematic consequences in the context of equality law, and it is important to provide clarity as to the various phases and intervening actors. Although it is certainly possible to further differentiate between different elements of decision making that form part of each particular stage, this threefold distinction is a convenient starting point for discussing algorithmic discrimination.

-
- 60 Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making: Opportunities and challenges* (Panel for the Future of Science and Technology (STOA) of the European Parliament) 5.
 - 61 Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies, Appendix, [B.5.6]. See further, also on the term 'emerging digital technologies', Expert Group on Liability and New Technologies – New Technologies Formation (2019), *Liability for Artificial Intelligence and other emerging digital technologies* (European Union) 11.
 - 62 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.2; Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 35 *Georgia State University Law Review* 1306, 1307. In a 2018 EU Communication, AI is defined as referring to 'systems that display intelligent behaviour by analysing their environment and taking action – with some degree of autonomy – to achieve specific goals'; European Commission (2018), Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions 'Coordinated Plan on Artificial Intelligence', COM(2018) 795 final (Brussels 2018) 1; idem in European Commission (2018), Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 'Artificial Intelligence for Europe', COM(2018) 237 final (Brussels 2018) 1. For another definition, see Hamon, R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* (JRC Technical Report, EU Science Hub) (European Commission, Brussels) 10.
 - 63 Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 35 *Georgia State University Law Review* 1306, 1308. See also Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* 19.
 - 64 Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 1306, 1322. See also Huq, AZ (2020), 'A Right to a Human Decision'.
 - 65 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 1.3.3.
 - 66 See also, based on a similar distinction of different stages in the algorithmic decision-making process, Xenidis, R and Senden, L (2020), 'EU Non-discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination', in Bernitz, U and others (eds), *General Principles of EU Law and the EU Digital Order* (Alphen aan den Rijn, Kluwer Law International) section 7.01[B]; Kleinberg, J and others (2019), 'Discrimination in the Age of Algorithms', available at: <https://arxiv.org/abs/1902.03731>.

1.3.1 Planning stage

The first stage distinguished by Kulk, Van Deursen and others is that of problem analysis and planning.⁶⁷ First of all, this encompasses defining a particular objective for the use of an algorithm by a company or public body. Such an objective may be to deploy personnel more efficiently, to set prices for certain products or services, to bring supply and demand together on the energy market, to detect social security fraud, to optimise risks calculations in relation to setting insurance premiums, to detect disinformation or hate speech on social media platforms, to help diagnose certain illnesses or diseases, and so on.⁶⁸

Once the relevant objectives have been established, they can be used to decide which type of algorithm is best suited to achieving each objective, if and how the algorithm can be fitted into pre-existing work flows and processes, and if and how it can or must be connected to other algorithms or automated processes. One of the factors that need to be taken into account in this decision-making process is how the output of the algorithm will be used. As explained in section 1.2, efficiency increasing automated decision making can sometimes require the use of rule-based algorithms, while self-learning algorithms can be more useful if the objective sets a need for predictions of human behaviour or profiling.⁶⁹ In making a choice it is also important to take account of the specific characteristics of an algorithm. Rule-based characteristics are typically highly predictable, in that all relevant parameters, variables and choices can be pre-determined as part of the development process. Once the algorithm is ready, there will therefore be no surprises. This also means that this type of algorithm is relatively rigid: it cannot independently take account of any changing contextual circumstances, such as new ideas on what would be an acceptable price or a reasonable fine. If it is believed that a rule-based algorithm is no longer generating 'good' decisions, it will have to be reprogrammed.

Similarly, supervised-learning algorithms are trained by using labelled and known set of data, which usually reflect the situation at a particular moment. This means that they can become outdated relatively quickly if contextual factors change (e.g. changing individual preferences and opinions or the developing use of particular expressions and wording in social media). Such supervised-learning algorithms therefore have to be updated and revalidated rather frequently and may be less suitable to operating in highly dynamic contexts.

In contrast, unsupervised-learning and deep-learning algorithms are much more adaptable and flexible: if the data changes because of societal developments, for example, these algorithms can train themselves to discover new patterns in the new data.⁷⁰ At the same time, the disadvantage of such adaptable algorithms is that it can be difficult to explain how they work and how they adapt, and users and even experts have little control over such adaptations.⁷¹

Finally, in the planning stage, users will need to make other important choices. For example, they have to decide whether they want to develop their own algorithm or purchase one that has already been built by an external provider (and possibly has been certified in conformity with given standards, such as the CE mark), or a combination of the two.⁷² Similarly, users have to make a decision on the datasets that will be used to build and train an algorithm (except when they have purchased a fully-developed, static algorithm that does not need any further development).⁷³ Sometimes such datasets are readily available

67 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 1.3.3.

68 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 1.3. See also Fagan, F and Levmore, S (2019) 'The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion' 93 *Southern California Law Review* 11.

69 On such choices, see e.g. Larus, J (2018) and others, 'When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making' (Technical Report Informatics Europe & EUACM).

70 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.2.

71 Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten* 49.

72 See Fundamental Rights Agency (FRA) (forthcoming), *AI and fundamental rights*.

73 See also Kleinberg, J and others (2019), 'Discrimination in the Age of Algorithms', available at: <https://arxiv.org/abs/1902.03731>.

from a company or public body or can be purchased from an external provider, but sometimes they still need to be constructed. Moreover, if existing datasets are used, there may be a need to supplement them with other data (e.g. specifically national data if a foreign dataset is bought).⁷⁴

1.3.2 Development stage

In the development stage, data scientists and other technology experts write the computer codes that are necessary to build the algorithm and, in the case of self-learning algorithms, allow them to be trained or engage in deep learning.⁷⁵ In addition, they have to decide how the algorithm can be connected to other pre-existing technical systems and applications in order to achieve the objectives set by the user.

If rule-based algorithms are to be built, the development stage comprises the unravelling of rules and decision-making processes into different steps and rebuilding them in computer code, making choices as to the appropriate variables and the type of decisions the algorithm eventually should make.⁷⁶ Technical experts often do this in cooperation with experts in the field, such as legal and policy experts, for systems that will assist in or take over administrative decision-making processes.

If a self-learning algorithm is to be used, part of the development process may be to decide exactly which type of learning has to be applied in order to achieve the objectives set in the planning stage. In addition, it must be decided which technologies for data analysis are best suited to achieving the set objectives, target variables have to be defined and valued and these have to be translated into computer-readable language and so on.⁷⁷ If self-learning algorithms are being developed, the development stage further comprises the preparation of data (e.g. labelling in supervised-learning processes, processing for ethical concerns, etc.) to make them suited for the process of training and learning. Finally, the development stage encompasses the actual training and feedback processes, and eventually the testing, validation and (if possible) certification of the algorithm.⁷⁸

1.3.3 Decision-making and use stage

Once an algorithm has been developed, tested and validated, it is ready to be used. This means that it can be fed with new input data and can start generating output that can be used to achieve the objectives set in the planning stage. Exactly how the algorithmic output is used can differ for different types of algorithms and will depend on the objectives of the user.

One possibility is that the algorithm directly generates a decision without further human intervention (*automated decision making* or ADM).⁷⁹ This may be relevant for routine decision-making processes, such as the imposition of fines in simple traffic offence cases or making bulk decisions in the areas of taxation or social security. In many such cases the decisions can be made using rule-based algorithms,⁸⁰ but self-learning algorithms may produce output that can also directly generate a decision and can be very

74 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 8.2.1. See also Fundamental Rights Agency (FRA) (forthcoming), *AI and fundamental rights*.

75 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 8.2.2.

76 Ibid.

77 Barocas, S and Selbst, A (2016), 'Big Data's Disparate Impact' 104 *California Law Review* 761, 768.

78 On these steps, see further e.g. Kroll, JA, Huey, J and others (2017), 'Accountable Algorithms' 165 *University of Pennsylvania Law Review* 633.

79 See further e.g. Larus, J (2018) and others, 'When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making' (Technical Report Informatics Europe & EUACM); Mendoza, I and Bygrave, LA (2017), 'The right not to be subject to automated decisions based on profiling' in Synodinou, T and others (eds), *EU Internet Law* (Berlin/Heidelberg, Springer). Article 22 GDPR contains a right not to be subjected to automated decision making based on profiling; on the GDPR, see further briefly section 1.5.1.2.

80 See further Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 35 *Georgia State University Law Review* 1306, 1317.

powerful in doing so.⁸¹ For example, many online platforms and web shops nowadays use self-learning algorithms for online price determination, making personalised offers to particular users, ensuring targeted and individualised newsfeeds or removing certain offensive or illegal posts.⁸² In addition, automated decisions can be made by means of a combination of different types of algorithms. A good example is self-driving cars, where many sensorial systems and different types of algorithms work together to allow the car to make the decision to brake or divert if it meets an obstacle.⁸³

The output of the algorithm may also be used to *support* a human decision, which means that there is a 'human in the loop' where the actual decision making is concerned.⁸⁴ In other words, there is 'teamwork' between humans and machines.⁸⁵ Examples of such teamwork include: a medical specialist might use an algorithmic analysis of medical imaging to check his diagnosis; a police unit manager might find support in predictive patterns of burglary risks in deciding to patrol more intensively in certain neighbourhoods; a social media employee might be assisted by algorithm-generated indications of instances of hate speech on a platform; a keen sportsperson might use an algorithmic tool that helps her to optimise her nutrient intake and training schedule; a bank might identify particular credit risks for an individual asking for a loan based on regression analyses; or a recruitment officer might make decisions on the suitability of job applicants supported by an algorithmic analysis of success factors. In all these cases, the human is the 'captain' of the 'team' and can ultimately ignore or overrule the decision suggested by the algorithm.⁸⁶

Finally, part of the use and decision-making stage is monitoring to see whether the algorithm continues to generate reliable and acceptable outcomes.⁸⁷ As mentioned above, rule-based and supervised-learning algorithms may easily become outdated, while complex machine-learning or deep-learning algorithms may unexpectedly generate unwarranted outcomes (e.g. discrimination) because they have learnt themselves to identify correlations that do not reflect causal relationships, or because they are not able to deal with certain types of new data.

1.4 Algorithmic characteristics and challenges

In section 1.2, different types of algorithms have been explained and attention has been paid to their functions and uses. Regardless of all their differences, algorithms have a number of commonalities that are important from a legal and ethical perspective, and so are discussed in this section. In addition, and directly related to these core characteristics, some general challenges are identified that algorithms present for the right to non-discrimination and equal treatment. The characteristics and challenges described in this section explain what makes algorithmic discrimination new and different from the types of 'human' discrimination that EU gender equality and non-discrimination law was originally crafted to address.⁸⁸ As such, they also help us to understand the legal, policy and conceptual questions that arise when considering algorithmic discrimination, which are central to Chapters 2 to 4 of the report.

-
- 81 Larus, J (2018) and others, 'When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making'.
 - 82 See e.g. Wachter, S (2020), 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' 35 *Berkeley Technology Law Journal*; Zuiderveen Borgesius, F and Poort, J (2017), 'Online Price Discrimination and EU Data Privacy Law' *Journal of Consumer Policy* 347.
 - 83 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, chapter 5.
 - 84 Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 1306, 1320.
 - 85 Fagan, F and Levmore, S (2019), 'The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion' 93 *Southern California Law Review*.
 - 86 Fagan, F and Levmore, S (2019), 'The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion' 93 *Southern California Law Review*, 6.
 - 87 See Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making: Opportunities and challenges* (Panel for the Future of Science and Technology (STOA) of the European Parliament) 73.
 - 88 Such differences and specificities are more generally cause for examining whether current EU legislation is able to address the specific risks of AI; see further European Commission (2020), 'White paper on artificial intelligence: a European approach to excellence and trust', COM(2020) 65 final (Brussels 2020) 10.

1.4.1 The human factor and the stereotyping and cognitive bias challenge

Even though algorithms operate increasingly autonomously, the role of human beings remains crucial for all of them.⁸⁹ This is obvious for rule-based algorithms, since human experts and computer programmers are directly responsible for deconstructing a decision-making process into different steps and translating them into computer instructions. The role of humans in machine-learning algorithms may be less obvious, but is still highly significant: humans are responsible for preparing and labelling the relevant training data and/or providing feedback to the algorithm in the iterative learning process.⁹⁰ Even in relation to deep-learning processes, which would seem to be relatively autonomous, humans play a role in developing the algorithms and checking the quality and validity of their output.⁹¹ Human beings are therefore of crucial importance in the development stage.⁹²

The role of humans in the planning and use stages, discussed in sections 1.3.1 and 1.3.3, is equally important. Humans decide on whether they want to develop and use an algorithm, what datasets they would like to use, and how they want to use the output of an algorithm.⁹³ In addition, humans are responsible for monitoring the continued quality and validity of the output of such algorithms, and for deciding to stop, replace or adapt algorithms that show dysfunctions.⁹⁴

Humans thus play a crucial and essential role in the programming, training and use of algorithms. Indeed, many people trust algorithms only if and because there is a 'human in the loop'.⁹⁵ At the same time, it is important to understand that the involvement of human beings bears particular risks in relation to algorithmic decision making, particularly from a perspective of equality and non-discrimination. It is well known that human reasoning shows flaws, biases, logical errors and fallacies, which may have an impact on the programming of algorithms.⁹⁶ Equally, the (personal, societal and therefore human-derived) data fed into algorithms in the training and use stages may also be non-neutral and biased, for instance because it reflects patterns of discrimination, as is further explained in section 1.4.2 below.⁹⁷ For that reason, the perpetuation of human bias has been typified as one of the key challenges for modern algorithmic societies:⁹⁸ algorithmic systems tend to simply 'reflect the values of their creators'.⁹⁹ Consequently, the mechanisms described here may easily lead to perpetuating prejudice, overbroad or harmful stereotypes and structural forms of discrimination. In other words, humans' discriminatory attitudes risk being translated and reflected in the algorithms that humans build.

89 See Katyal, SK (2019), 'Private accountability in the age of artificial intelligence' 66 *UCLA Law Review* 54, 62; Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* 25; Fagan, F and Levmore, S (2019), 'The impact of artificial intelligence on rules, standards, and judicial discretion' 93 *Southern California Law Review* 11; Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten*, 48-9.

90 Katyal, SK (2019), 'Private accountability in the age of artificial intelligence', 54, 67; Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law' 55 *Common Market Law Review* 1143, 1147; Crawford, K (2013), 'The Hidden Biases in Big Data' *Harvard Business Review*, available at: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>.

91 Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* 26.

92 See above, section 1.3.2.

93 Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* 26; Katyal, SK (2019), 'Private accountability in the age of artificial intelligence' 66 *UCLA Law Review* 54, 67.

94 Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* 27.

95 See European Commission (2020), 'White Paper on Artificial Intelligence', 21; see further also Independent High Level Expert Group on Artificial Intelligence (AIHLEG) (2019), 'Ethics Guidelines for Trustworthy AI' (Brussels) 4, 12.

96 Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* 20; Fagan, F and Levmore, S (2019), 'The impact of artificial intelligence on rules, standards, and judicial discretion' 93 *Southern California Law Review* 11, 14; Kleinberg, J and others (2019), 'Discrimination in the age of algorithms', available at: <https://arxiv.org/abs/1902.03731>, 11; Katyal, SK (2019), 'Private accountability in the age of artificial intelligence' 66 *UCLA Law Review* 54, 66; Mittelstadt, B (2016), 'Auditing for transparency in content personalization systems' 10 *International Journal of Communication* 4991, 4993. On the notion of 'bias', see more specifically section 1.5.1.1.

97 See e.g. Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making: Opportunities and challenges* (Panel for the Future of Science and Technology (STOA) of the European Parliament) 40; Schermer, B (2011), 'The limits of privacy in automated profiling and data mining' 27 *Computer Law & Security Review* 45.

98 Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law' 55 *Common Market Law Review* 1143, 1144.

99 Crawford, K and Whittaker, M (2016), *The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*, available at: https://ainowinstitute.org/AI_Now_2016_Report.pdf, 9.

Another particular challenge arises from cognitive biases that are at play when humans are assisted by algorithms.¹⁰⁰ It has been shown, for example, that human decision makers will trust the outcomes of the algorithm, being convinced that the algorithm probably ‘knows’ or performs better than they would do.¹⁰¹ This so-called ‘automation bias’ in favour of the algorithm may lead to ‘commission errors’ or rubber-stamping: trusting the quality and authority of the algorithm, human decision makers tend to embrace the decision it suggests.¹⁰² If a human decision maker wants to take a more critical stand and disagrees with the suggested outcome, she might feel an additional pressure to motivate her decision to deviate from the computer output.¹⁰³ This may not be easy, even if the decision maker’s own intuition and experience inform her that a certain decision simply cannot be right.¹⁰⁴ Similarly, algorithmic output may lead to ‘anchoring’, for example in imposing sanctions by judges.¹⁰⁵ Based on algorithmic analysis of large amounts of previous cases, an application might suggest that in a particular case of shoplifting or burglary, a particular fine or prison sentence would be indicated. This then involuntarily forms an anchor for the judge, who will tend to stay relatively close to the indicated level of the sanction, even if she might have arrived at a very different sanction had she been able to make a decision fully on her own.¹⁰⁶ When combined with the risks of bias in data or flaws in the programming of an algorithm, these cognitive phenomena of rubber-stamping, automation bias and anchoring pose an additional challenge in terms of non-discrimination.

In sum, as Kleinberg and others have aptly remarked, ‘the Achilles’ heel of all algorithms is the humans who build them and the choices they make’.¹⁰⁷ A critical element of regulating algorithms is therefore ‘regulating humans’.¹⁰⁸ This is what this report calls the *stereotyping and cognitive bias challenge*.

1.4.2 The data challenge

As has been illustrated above, the quality, accuracy, validity and reliability of algorithms depend on the quality of the input provided.¹⁰⁹ Several problems may arise in this respect.

-
- 100 See Tversky, A and Kahneman, D (1974), ‘Judgment under uncertainty: Heuristics and biases’ 185 *Science* (4157) 1124. For a more recent review, see Katyal, SK (2019), ‘Private Accountability in the Age of Artificial Intelligence’ 66 *UCLA Law Review* 54, 79. See also Lieder, F and others (2018), ‘The anchoring bias reflects rational use of cognitive resources’ 25 *Psychonomic Bulletin & Review* 322, 322-3.
- 101 See e.g. Goddard, K and others (2012), ‘Automation bias: a systemic overview of frequency, effect mediators, and mitigators’ 19 *Journal of the American Medical Informatics Association* 121; Skitka, LJ, Mosier, KL and Burdick, M (1999), ‘Does automation bias decision-making?’ 51 *International Journal of Human-Computer Studies* 991; Parasuraman, R and Riley, V (1997), ‘Humans and automation: use, misuse, disuse, abuse’ 39 *Human Factors* 230.
- 102 See e.g. Goddard, K and others (2012), ‘Automation bias: a systemic overview of frequency, effect mediators, and mitigators’ 19 *Journal of the American Medical Informatics Association* 121; Skitka, LJ, Mosier, KL and Burdick, M (1999), ‘Does automation bias decision-making?’ 51 *International Journal of Human-Computer Studies* 991, 993; Parasuraman, R and Riley, V (1997), ‘Humans and automation: use, misuse, disuse, abuse’ 39 *Human Factors* 230; see also Katyal, SK (2019), ‘Private accountability in the age of artificial intelligence’ 66 *UCLA Law Review* 54, 73; Doshi-Velez, F, Korz, M and others (2017), ‘Accountability of AI under the law: the role of explanation’ (Berkman Klein Center for Internet & Society working paper), available at: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372584>, 2. On a variation of this bias (classification bias) see further Kim, PT (2017), ‘Data-Driven Discrimination at Work’ 58 *William and Mary Law Review* 857, 866. See critically Huq, AZ (2020), ‘A Right to a Human Decision’ 105 *Virginia Law Review* 50-1.
- 103 Skitka, LJ, Mosier, KL and Burdick, M (1999), ‘Does automation bias decision-making?’ 991, 1003.
- 104 Goddard, K and others (2012), ‘Automation bias: a systemic overview of frequency, effect mediators, and mitigators’ 121, 124; Skitka, LJ, Mosier, KL and Burdick, M (1999), ‘Does automation bias decision-making?’ 991, 993.
- 105 See Tversky, A and Kahneman, D (1974) ‘Judgment under uncertainty: Heuristics and biases’ 185 *Science* (4157), 1228-9; Nisbett, RE and Ross, L (1980) *Human inference: Strategies and shortcomings of social judgment* (Prentice Hall, Englewood Cliffs). For a fuller and more recent review, see Lieder, F and others (2018), ‘The anchoring bias reflects rational use of cognitive resources’ 25 *Psychonomic Bulletin & Review* 322, 323.
- 106 See Tversky, A and Kahneman, D (1974) ‘Judgment under uncertainty: Heuristics and biases’, 1228-9; Nisbett, RE and Ross, L (1980) *Human inference: Strategies and shortcomings of social judgment*. For a fuller and more recent review, see Lieder, F and others (2018), ‘The anchoring bias reflects rational use of cognitive resources’, 322, 323.
- 107 Kleinberg, J and others (2019), ‘Discrimination in the age of algorithms’, available at: <https://arxiv.org/abs/1902.03731>, 4. On the limited ‘objectivity’ of algorithms, see also Gillespie, T (2014), ‘The Relevance of Algorithms’ in Gillespie, T, Boczkowski, PJ and Foot, KA (eds), *Media technologies: Essays on communication, materiality, and society* (Cambridge, MIT Press) 167, 181
- 108 Ibid.
- 109 See Katyal, SK (2019), ‘Private accountability in the age of artificial intelligence’ 66 *UCLA Law Review* 54, 68; Information Commissioner’s Office (ICO) (2017), *Big data, artificial intelligence, machine learning and data protection*, available at: <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>, paras 94 et seq.

First, an otherwise correctly designed algorithmic system may be fed with incorrect information. To give a simple example: if a sensor dysfunction wrongly leads to detecting the violation of a traffic rule,¹¹⁰ a rule-based algorithm will not be able to identify the mistake itself and will simply generate a decision to impose a fine based on the assumption that a traffic offence has been committed. However, humans would consider that outcome to be unfair, because the car driver is being fined for an offence that she has not committed.¹¹¹

Another inaccuracy that may occur is that the data used to train an algorithm is unrepresentative of the general population, inadequately deals with outliers or does not include particular minority groups.¹¹² This type of inaccuracy may easily lead to discrimination, since this may result in an algorithm automatically reflecting the imbalanced data on which it has been trained. The classic example is that of the facial recognition software that performs less well on Black women's faces than on White women's and Black men's faces because Black women are under-represented in the dataset used to train the algorithm.¹¹³

Finally, and very relevant from a non-discrimination perspective, if a self-learning algorithm is fed with unbalanced or biased data it is very likely to generate equally unbalanced and biased output based on its detection of correlations and patterns in that data.¹¹⁴ As Barocas and Selbst have explained, 'approached without care, data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society'.¹¹⁵ Often this is summarised as 'rubbish in, rubbish out' or 'bias in, bias out'.¹¹⁶ As was noted in the *AI Now* report from 2016, 'there is the risk that AI systems trained on this data will produce models that replicate and magnify those biases. In such cases, AI systems would exacerbate the discriminatory dynamics that create social inequality, and would likely do so in ways that would be less obvious than human prejudice and implicit bias'.¹¹⁷ This is exacerbated if a self-learning algorithm is using the output it has generated based on flawed data to further 'improve' itself. In that case a feedback loop can be created that reinforces already existing patterns of structural discrimination by 'reifying' and further enacting discriminatory correlations.¹¹⁸

Because of the human factor, discussed in section 1.4.1 above, such data problems are very likely to occur in the planning and development stage, for example when selecting the data that is to be used or when preparing or labelling the data. Clearly it is imperative to be aware of the quality, accuracy and reliability

-
- 110 Crawford, K and Whittaker, M (2016), *The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*, available at: https://ainowinstitute.org/AI_Now_2016_Report.pdf, 6; compare also Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 35 *Georgia State University Law Review* 1306, 1316. Feeding false data to an algorithm may sometimes have even been done intentionally as 'data poisoning' – see Hamon, R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* (JRC Technical Report, EU Science Hub) (European Commission, Brussels) 16.
- 111 For more examples, see Barocas, S and Selbst, A (2016), 'Big Data's Disparate Impact' 104 *California Law Review* 761, 684.
- 112 Katyal, SK (2019), 'Private accountability in the age of artificial intelligence', 54, 70; Barocas, S and Selbst, A (2016), 'Big Data's Disparate Impact', 761, 687.
- 113 See Buolamwini, J and Gebru, T, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification* (Proceedings of Machine Learning Research 2018).
- 114 cf Katyal, SK (2019) 'Private accountability in the age of artificial intelligence', 54, 66; Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law' 55 *Common Market Law Review* 1143, 1146; Information Commissioner's Office (ICO) (2017), *Big data, artificial intelligence, machine learning and data protection*, paras 96ff; Crawford, K and Whittaker, M (2016), *The AI Now Report*, 6; Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making*, 40; Schermer, B (2011), 'The limits of privacy in automated profiling and data mining' 27 *Computer Law & Security Review* 45.
- 115 Barocas, S and Selbst, A (2016), 'Big Data's Disparate Impact' 104 *California Law Review* 761, 674.
- 116 Barocas, S and Selbst, A (2016), 'Big Data's Disparate Impact' 761, 683; ICO (2017), *Big data, artificial intelligence, machine learning and data protection*, para 96.
- 117 Crawford, K and Whittaker, M (2016), *The AI Now Report*, 6-7. See also Xenidis, R and Senden, L (2020), 'EU non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination', in Bernitz, U and others (eds), *General Principles of EU Law and the EU Digital Order* (Alphen aan den Rijn, Kluwer Law International) section 7.01[C].
- 118 Xenidis, R and Senden, L (2020), 'EU Non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination', section 7.01[C][2]. See also Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies, Appendix [5].

of the data used for labelling, training, feedback and learning. In this report, this is described as the *data challenge*.¹¹⁹

1.4.3 The correlation and proxies challenge

Algorithms that are used for pattern recognition (see section 1.2.2.1) are often very good at detecting correlations and patterns in large volumes of data.¹²⁰ However, correlations do not always correspond to causal relationships.¹²¹ For example, gender might negatively correlate with level of performance at work, not because of a causal relationship, but because women historically have been consistently evaluated more negatively than men for the same work performance.¹²² This example shows that decisions based on correlations found by an algorithm may not always be acceptable from a human perspective, since human thinking is informed by normative or ethical considerations and causation logic.¹²³ Moreover, algorithms may reproduce and strengthen existing patterns of inequality by reifying discriminatory correlations.

This *correlation challenge* is exacerbated by the fact that algorithms are very good at detecting ‘proxies’.¹²⁴ For example, algorithms may be trained not to base an output on certain personal characteristics such as gender, ethnic origin or religion to avoid discrimination. Nevertheless, they may easily detect other variables and ‘neutral’ data points that are very closely related to those characteristics, ranging from certain types of clicking behaviour to zip codes and preferences for particular types or colours of cars.¹²⁵ If algorithms take account of such ‘proxy variables’ in identifying correlations, they can approach the original prohibited characteristic very closely, with the same discriminatory outcomes but without this being highly visible.¹²⁶ This can be coincidental or a result of deeply engrained, structural discrimination,¹²⁷ but it can also be intentional, which is known as ‘masking’: a trivial and non-suspect proxy is used to mask a case of conscious discrimination based on a protected ground.¹²⁸ This also makes clear that simply omitting certain personal data in the process of developing an algorithm, such as information

-
- 119 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.2; ICO (2017), *Big data, artificial intelligence, machine learning and data protection*, para 96; Crawford, K (2013) ‘The Hidden Biases in Big Data’. See also European Commission (2020), ‘White Paper on Artificial Intelligence’, 18-9.
- 120 See further Williams, BA, Brooks, CF and Shmargad, Y (2018), ‘How algorithms discriminate based on data they lack: challenges, solutions, and policy implications’ 8 *Journal of Information Policy* 78, 83.
- 121 Schölkopf, B (2019), ‘Causality for Machine Learning’, available at: <https://arxiv.org/pdf/1911.10500.pdf>; Kleinberg, J and others (2019), ‘Discrimination in the Age of Algorithms’, available at: <https://arxiv.org/abs/1902.03731>, 20; Katyal, SK (2019), ‘Private accountability in the age of artificial intelligence’ 66 *UCLA Law Review* 54, 71; Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten*, 50; Schermer, B (2011), ‘The limits of privacy in automated profiling and data mining’ 27 *Computer Law & Security Review* 45.
- 122 See e.g. Kullmann, M (2018), ‘Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law’; Kim, PT (2017), ‘Data-Driven Discrimination at Work’ 58 *William and Mary Law Review* 857. See also Wachter, S (2020), ‘Affinity profiling and discrimination by association in online behavioural advertising’ 35 *Berkeley Technology Law Journal* 19.
- 123 Kim, PT (2017), ‘Data-Driven Discrimination at Work’, 857, 881; cf also ICO (2017), *Big data, artificial intelligence, machine learning and data protection*, para 117.
- 124 See further e.g. Gillis, TB and Spiess, JL (2019), ‘Big Data and Discrimination’ 86 *University of Chicago Law Review* 459, 468ff; Barocas, S and Selbst, A (2016), ‘Big Data’s Disparate Impact’ 104 *California Law Review* 761, 691.
- 125 E.g. Gillis, TB and Spiess, JL (2019), ‘Big Data and Discrimination’, 459, 469; Wachter, S (2020), ‘Affinity profiling and discrimination by association in online behavioural advertising’, 17; Kullmann, M (2018), ‘Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law’, p. 7; Williams, BA, Brooks, CF and Shmargad, Y (2018), ‘How algorithms discriminate based on data they lack: challenges, solutions, and policy implications’ 8 *Journal of Information Policy* 78, 84, 89; Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’ 55 *Common Market Law Review* 1143, 1148; Kim, PT (2017), ‘Data-Driven Discrimination at Work’, 857, 877.
- 126 Xenidis, R and Senden, L (2020), ‘EU Non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination’, in Bernitz, U and others (eds), *General Principles of EU Law and the EU Digital Order* (Alphen aan den Rijn, Kluwer Law International) section 7.01[C][2]; Barocas, S and Selbst, A (2016), ‘Big Data’s Disparate Impact’ 104 *California Law Review* 761, 691.
- 127 On the structural causes of proxy discrimination, see in particular Xenidis, R and Senden, L (2020), ‘EU Non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination’, section 7.01[C][2] and Williams, BA, Brooks, CF and Shmargad, Y (2018), ‘How algorithms discriminate based on data they lack: challenges, solutions, and policy implications’, 87-8.
- 128 Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’, 1143, 1149; Kim, PT (2017), ‘Data-Driven Discrimination at Work’, 857, 884; Barocas, S and Selbst, A (2016), ‘Big Data’s Disparate Impact’, 761, 692; Žliobaitė, I and Custers, BHM (2016), ‘Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models’ 24 *Artificial Intelligence and Law* 183.

about gender or ethnicity, does not guarantee that discrimination is avoided.¹²⁹ Although that may help to reduce the possibility of 'overt' (as opposed to 'covert'¹³⁰) direct discrimination, due to the prevalence of proxies there may still be room for indirect discrimination.¹³¹ This is what is called the *proxies challenge*.

1.4.4 The transparency and explainability challenge

Another common characteristic of (and challenge related to) algorithms is that they are opaque and difficult to explain, especially to non-experts.¹³² Even relatively straightforward, rule-based algorithms may be so complex that outsiders cannot easily comprehend their workings.¹³³ It is even more difficult to understand for people exactly how self-learning algorithms work, in particular deep-learning algorithms.¹³⁴ Such algorithms might still be transparent to technical experts, especially if they are given all the necessary information on the relevant source codes, input variables, parameters and threshold values,¹³⁵ but lay people will find it very difficult to understand how an individual risk or a specific pattern is identified by means of a self-learning algorithmic application.¹³⁶ Obviously, this will be even more true of intricately interconnected sets of algorithms that function with some degree of autonomy and can almost mimic human intelligence, as may be the case for AI systems.

The lack of transparency for outsiders and lay people, combined with the difficulties of explaining the workings of an algorithm make it difficult for human decision makers to identify any of the flaws, biases or ill-qualified correlations that may be part of the algorithmic process.¹³⁷ Many people who are subjected to algorithmic decision making will never know exactly how the decisions that affect them on a daily basis are made, whether they relate to price-setting or an employment offer and whether they influence their insurance premiums or lead to the removal of their social media posts. This opacity and lack of information makes discrimination and bias difficult to discover.¹³⁸ Hence, in the absence of algorithmic transparency and explainability (a process by which the 'black box' of an algorithm is made intelligible

-
- 129 Kroll, JA, Huey, J and others (2017), 'Accountable Algorithms' 165 *University of Pennsylvania Law Review* 633, 681; Žliobaitė, I and Custers, BHM (2016), 'Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models' 24 *Artificial Intelligence and Law* 183.
- 130 See eg Besson, S (2008), 'Gender Discrimination under EU and ECHR Law: Never Shall the Twain Meet?' 8 *Human Rights Law Review* (4) 647 and Arnardóttir, OM (2003), *Equality and non-discrimination under the European Convention on Human Rights* (Martinus Nijhoff Publishers).
- 131 Wachter, S (2020), 'Affinity profiling and discrimination by association in online behavioural advertising' 35 *Berkeley Technology Law Journal* 17.
- 132 See further e.g. Burrell, J (2016), 'How the machine "thinks": Understanding opacity in machine learning algorithms' 3 *Big Data & Society*. On the difference between explainability and transparency, see e.g. Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making: Opportunities and challenges* (Panel for the Future of Science and Technology (STOA) of the European Parliament) 27. Other scholars have more faith in the transparency and explainability of algorithms and even see the potential of explainability as an asset in fighting discrimination; see e.g. Kleinberg, J and others (2019), 'Discrimination in the Age of Algorithms', available at: <https://arxiv.org/abs/1902.03731>.
- 133 Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* 21-2; Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.1.1; Mittelstadt, B (2016), 'Auditing for transparency in content personalization systems' 10 *International Journal of Communication* 4991, 4996; Van Otterlo, M (2013), 'A machine learning view on profiling', in Hildebrandt, M and De Vries, K (eds), *Privacy, due process and the computational turn: Philosophers of law meet philosophers of technology* (London, Routledge) 46.
- 134 Gillis, TB and Spiess, JL (2019), 'Big Data and Discrimination' 86 *University of Chicago Law Review* 459, 475; Larus, J and others (2018), 'When computers decide: European recommendations on machine-learned automated decision making' (Technical Report Informatics Europe & EUACM) 9; Mittelstadt, B (2016), 'Auditing for transparency in content personalization systems' 10 *International Journal of Communication* 4991, 4997.
- 135 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.3. Even then, however, it can be difficult to detect the variables that self-learning algorithms rely on; see Gillis, TB and Spiess, JL (2019), 'Big Data and Discrimination', 459, 476; Kroll, JA, Huey, J and others (2017), 'Accountable Algorithms' 165 *University of Pennsylvania Law Review* 633, 657.
- 136 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, section 2.3; Kim, PT (2017), 'Data-Driven Discrimination at Work' 58 *William and Mary Law Review* 857, 881.
- 137 See Kroll, JA, Huey, J and others (2017), 'Accountable Algorithms' 165 *University of Pennsylvania Law Review* 633. See also European Commission (2020), 'White Paper on Artificial Intelligence', 12.
- 138 See e.g. Kim, PT (2017), 'Data-Driven Discrimination at Work' 58 *William and Mary Law Review* 857.

and understandable to human experts), it becomes a challenge for potential victims of discrimination, as well as monitoring and supervisory bodies and courts, to detect and provide evidence of discrimination.¹³⁹

1.4.5 *The scale and speed challenge*

As explained in section 1.3, algorithmic decision making can be used to allow for high-speed, sometimes automated decision making on a very large scale.¹⁴⁰ This can certainly help to enable extensive decision making in the public sector, for example in relation to traffic fines, routine social security and taxation decisions or the granting of permits. In the private sector, algorithms are also increasingly used to automatise decision making. Well-known examples are price-setting by web shops based on individual preferences and buying behaviour, making individual offers by platforms such as Uber or Airbnb, or individual targeting by video platforms or newsfeeds. Different types of algorithms can be used to effectuate such volumes and speed. Sometimes rule-based algorithms work best, while machine-learning or deep-learning algorithms are better at generating the desired effects in other cases. Nevertheless, what all algorithms have in common is that they allow for decision making on a much larger scale than traditional human decision making is capable of, and with unprecedented speed.¹⁴¹

This innovation in the scale and speed of decision making poses a new and general challenge from the perspective of non-discrimination. This is even more true if we take account of the characteristics and challenges discussed above.¹⁴² Flaws in human thinking, the existence of structural forms of societal discrimination and stereotyping, and the lack of representative and accurate data may negatively influence the process of designing, developing and using an algorithm. If insufficient checks are made, the particular characteristics of algorithmic decision making can cause algorithmic discrimination to 'spread' at a wider scale and a much quicker pace than 'human' discrimination could do.

1.4.6 *The responsibility challenge*

Finally, it is important to note that a variety of different players are involved in the stages of algorithmic decision making discussed in section 1.3. Different people or companies are responsible for setting the objectives, deconstructing decision-making processes, programming and training algorithms, collecting and preparing the training data, using algorithms for decision making, monitoring and supervising their effect, and so on.¹⁴³ Consequently, if at some point a discriminatory outcome is detected (for instance, because an algorithm systematically suggests that men should be promoted to a certain position rather than women), it may be very difficult for the victim of discrimination or for supervisory or monitoring bodies to know whom to hold responsible, liable and/or accountable for that discriminatory outcome among the various players involved (the developers, the sellers or the end user (in the example above,

139 See further Xenidis, R and Senden, L (2020), 'EU non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination', in Bernitz, U and others (eds), *General Principles of EU Law and the EU Digital Order* (Alphen aan den Rijn, Kluwer Law International) section 7.03[A]. For this reason, it is often suggested that measures should be taken to improve explainability; see e.g. Independent High-Level Expert Group on Artificial Intelligence (AIHLEG) (2019), 'Ethics Guidelines for Trustworthy AI' (Brussels), 13. On explainability, see further section 1.4.4.

140 Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* 21. Cf also European Commission (2020), 'White Paper on Artificial Intelligence: A European approach to excellence and trust', COM(2020) 65 final (Brussels 2020) 11.

141 See in more detail eg Kim, PT (2017), 'Data-Driven Discrimination at Work' 58 *William and Mary Law Review* 857, 861. See also European Commission, 'White Paper on Artificial Intelligence', 11.

142 See in particular Kim, PT (2017), 'Data-Driven Discrimination at Work', 857, 861.

143 See Rammert, W (2008), 'Where the action is: Distributed agency between humans, machines, and programs', Technical University Technology Studies Working Papers, TUTS-WP-4-2008. See also Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies, Appendix, para 14.

the HR service) of the algorithm).¹⁴⁴ This is even more true in situations where different algorithms and enabling technologies work together, as is often the case in AI applications.¹⁴⁵ Identifying the ‘agent’ (person, body, institution, technological application or company) responsible for a case of discrimination therefore poses a particular challenge in relation to algorithms.¹⁴⁶

1.5 Terminology and interactions between gender equality and non-discrimination law and data protection law

1.5.1 Terminology: ‘bias’ and ‘fairness’ versus ‘discrimination’ and ‘equality’

‘Bias’ and ‘fairness’ are important notions surrounding the discussions on algorithmic discrimination.¹⁴⁷ While the meaning of these terms clearly overlaps with that of the legal notions of discrimination and equality, their relationship is not completely evident and needs careful articulation. Because discussions about algorithms, (semi)autonomous systems and discrimination have been expanding from computer science to other disciplines including law, core concepts need to be adapted to better reflect disciplinary idiosyncrasies. The notions of ‘bias’ and ‘fairness’ are grounded in statistics and ethics and have specific meanings that are not necessarily well-suited to capturing the specific problems that arise in relation to the law. This section offers clarifications on the terminology used throughout this report and its relationship with closely related notions and concepts.

The terms ‘bias’ and ‘fairness’ have historically been used by computer scientists to describe a range of ethical problems linked to the operation and outcome of algorithms. Specifically, ‘algorithmic bias’ refers to ‘a systematic error’ of any kind in the outcome of algorithmic operations.¹⁴⁸ Bias therefore has a much wider meaning than discrimination as it is not only concerned with unfair errors but with all kinds of ‘systematic’ errors, which can include those of a statistical, cognitive, societal, structural or institutional nature. When invoked in the particular context of ‘fairness’, however, ‘algorithmic bias’ refers to a particular type of error that ‘places privileged groups at a systematic advantage and unprivileged groups at a systematic disadvantage’.¹⁴⁹ This definition shares commonalities with the legal definition of discrimination understood as the differential unfavourable treatment of an individual or group or the disproportionately disadvantageous impact of a given measure or policy on a specific group. However, the term ‘algorithmic bias’ is more encompassing than the legal term ‘algorithmic discrimination’ as it refers to any kind of disadvantage that could be viewed as ethically or morally wrong. For example, an algorithm that disadvantages low-income groups and privileges people with high incomes could be seen as entailing a form of algorithmic bias from an ethical point of view.¹⁵⁰ From a legal point of view, however, algorithmic discrimination only pertains to the unjustified unfavourable treatment of, or disadvantage experienced by, specific categories of population protected by the law either explicitly (e.g. protected grounds) or implicitly (e.g. general or open-textured non-discrimination clauses). For example, in the context of EU gender

144 Expert Group on Liability and New Technologies – New Technologies Formation (2019), *Liability for Artificial Intelligence and other emerging digital technologies* (European Union) 20; Xenidis, R and Senden, L (2020), ‘EU non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination’, section 7.03[A]; Schermer, B (2011) ‘The limits of privacy in automated profiling and data mining’ 27 *Computer Law & Security Review* 45; see also European Commission (2020), ‘White Paper on Artificial Intelligence’, 12. For that reason, it has been argued that one person should be identified as responsible for the output of algorithms as one of the principles of accountable algorithms; see Diakopoulos, N and Friedler, S (2016), ‘How to Hold Algorithms Accountable?’ *MIT Technology Review*, available at: <https://www.technologyreview.com/2016/11/17/155957/how-to-hold-algorithms-accountable/>.

145 Expert Group on Liability and New Technologies – New Technologies Formation (2019), *Liability for Artificial Intelligence and other emerging digital technologies* (European Union) 20.

146 See Rammert, W (2008), ‘Where the action is: Distributed agency between humans, machines, and programs’, Technical University Technology Studies Working Papers, TUTS-WP-4-2008.

147 On the impact of bias on algorithms, see also sections 1.4.1 and 1.4.2.

148 Bellamy, RKE and others (2018), ‘AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias’, available at: arXiv:181001943 [csAI].

149 Bellamy, RKE and others (2018), ‘AI Fairness 360’.

150 The same could be said of an algorithm that systematically disadvantages people in (involuntary) long-term unemployment, which was one of the criteria at stake in the controversy around the **Austrian** AMS algorithm, see sections 3.1.2.1. and 3.3.2.1.

equality and non-discrimination law, algorithmic discrimination refers to discrimination based on one of the six grounds explicitly listed in and protected under Article 19 TFEU, that is sex, race or ethnic origin, disability, sexual orientation, religion or belief and age. This is why the term ‘algorithmic discrimination’ will be used throughout this report to refer to the types of algorithmic bias that are problematic from the point of view of EU gender equality and non-discrimination law.

Similarly, the notion of ‘algorithmic fairness’ has been traditionally used by computer scientists to describe a set of procedures aiming at avoiding bias so as to ensure outcomes that respect given ethical standards. Algorithmic fairness is a multi-faceted concept and computer scientists rely on the various types of fairness delineated by ethicists and philosophers, for example group, individual, procedural, outcome-based, counter-factual fairness, etc. As Kirkpatrick explains, ‘[f]airness is not necessarily clean-cut, given the competing interests, whether looking at commercial interests (profit versus access to products and services) or within the justice system, which must balance public safety, administrative efficiency, and the rights of defendants’.¹⁵¹ The notion of ‘fairness’ is therefore grounded in moral and ethical principles, the meaning of which can vary contextually. From this perspective, it is clear that ‘fairness’ overlaps with, but also goes beyond the legal definition of equality or non-discrimination. While the equality model underpinning non-discrimination law has been the subject of intense scholarly discussion in various contexts, it is safe to assume that in the context of EU law, the principle of equality has a more restrictive scope than the concept of ‘fairness’ as understood by computer scientists. The equality principle in EU law indeed only imposes obligations regarding the protection of the six ‘grounds’ mentioned above. The EU equality principle is also negative in the sense that it broadly covers a prohibition to treat or impact in an unfavourable manner, whereas algorithmic fairness procedures might be much more wide-ranging and go beyond such a conception.¹⁵²

Throughout this report, the terms ‘equality’ and ‘non-discrimination’ will be used to refer to the legal principles underpinning strategies against ‘algorithmic discrimination’ in the context of EU law as they allow for a more precise pinning down of legal obligations than the broad ethics-oriented concept of ‘algorithmic fairness’. In addition, it seems important to stress that the notions of ‘bias’ and ‘fairness’ that structure the discourse on algorithmic discrimination need to be tuned to the legal framework that protects the principle of equality in EU law. That is why Chapter 2 raises the questions of whether the EU legal framework adequately captures algorithmic discrimination, how algorithmic discrimination challenges this legal framework and where potential frictions and inadequacies arise.

1.5.2 Interactions between non-discrimination law and data protection

Scholarly discussions have drawn attention to the role of data protection law and legal obligations related to privacy in the prevention of algorithmic discrimination. The rationale is that certain categories of data – for instance race, gender, sexual orientation, etc. – are particularly sensitive because they can easily lead to unlawful discrimination if processed without particular precautions. This is reflected in the EU’s General Data Protection Regulation (GDPR), which identifies ‘special categories of personal data’ or ‘sensitive data’.¹⁵³ The Regulation recognises that ‘[p]ersonal data which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms merit specific protection as the context of their processing could create significant risks to the fundamental rights and freedoms’.¹⁵⁴ In particular, Recital 71 of the GDPR indicates that ‘[i]n order to ensure fair and transparent processing in respect of the data subject [...] the [data] controller should [...] prevent, inter alia, discriminatory effects on natural persons on

151 Kirkpatrick, K (2016), ‘Battling algorithmic bias: how do we ensure algorithms treat us fairly?’ 59 *Communications of the ACM*.

152 See e.g. the three types of fairness-enhancing mechanisms delineated in Pessach, D and Shmueli, E (2020), ‘Algorithmic Fairness’, available at: [arXiv:200109784v1](https://arxiv.org/abs/200109784v1) [csCY]. In EU law, positive action is allowed but there is no legal obligation for Member States to adopt such measures.

153 Recital 10 of Regulation 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1.

154 Recital 51 of the General Data Protection Regulation.

the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or processing that results in measures having such an effect'. Here the interaction between data protection law and non-discrimination law is evident. Nevertheless, this interaction poses a number of questions as the provisions do not neatly overlap.

The list of categories of data the processing of which could give risk to discrimination does not neatly fit with the list of protected grounds under EU gender equality and non-discrimination law. Importantly, the issue of gender equality or sex discrimination is altogether absent from the GDPR and neither gender nor sex are mentioned as sensitive categories of personal data. Racial or ethnic origin, religion or belief and sexual orientation are explicitly mentioned both in relation to discrimination in Recital 71 and in relation to the prohibition of processing such data, but the recital does not refer to 'sex' or grounds such as 'age' and 'disability'. Similarly, Article 9(1) GDPR prohibits the 'processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation'.¹⁵⁵ While the list is much broader than Article 19 TFEU on the prohibition of discrimination, it does not explicitly mention 'sex', 'disability' and 'age' either. It might be inferred that 'disability' is understood to be included in the terms 'health status' or 'data concerning health' but 'age' and 'sex', as protected grounds, are more difficult to read in these two provisions of the GDPR. The absence of an outright prohibition on processing such data categories could be justified by a variety of reasons pertaining to the possibilities for legitimate and non-discriminatory use of this data (although Article 9(2), 9(3) and 9(4) provides an explicit list of derogations concerning the legitimate use of such data), yet the absence of any mention of 'gender equality', 'age discrimination' and 'disability discrimination' in Recital 71 concerning discriminatory risks is more difficult to understand.

The approach taken by EU data protection law, and in particular the GDPR, towards preventing discrimination is furthermore substantially different from that of non-discrimination law and pertains to the degree of automation of the processing of sensitive data. The GDPR considers the presence of a human in the loop as a form of preventive safeguard, as Recital 71 makes clear:

'The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention. Such processing includes 'profiling' that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her.'

However, there is no clarity concerning the form of human supervision involved. This is problematic in view of the existence of so-called automation biases, as explained above.¹⁵⁶ In fact, prohibiting full automation does not ensure the absence of discrimination. In addition, Article 5(1) of the GDPR clarifies that some of the principles underpinning the processing of personal data are 'lawfulness, fairness and transparency', while Article 5(2) mentions the principle of 'accountability'. By contrast, the notion of 'discrimination' is only mentioned three times to describe the risks posed by the processing of sensitive personal data.¹⁵⁷ In turn, 'equality' is only mentioned twice in relation to processing data in the context of employment.¹⁵⁸ As a result, the concepts on which the GDPR relies in relation to the question of algorithmic discrimination

155 Recital 51 of the General Data Protection Regulation.

156 See section 1.4.1.

157 Recitals 71, 75, 85 of the General Data Protection Regulation.

158 Recital 155 and Article 88 of the General Data Protection Regulation.

are quite different from those central to gender equality and non-discrimination law and the link between the two areas is not made explicit by the GDPR. The approach taken by the GDPR to sensitive data offers some guarantees regarding some of the protected grounds covered by EU non-discrimination law, but also evidences gaps, not least in relation to the protection of gender equality.

Despite different conceptual approaches to the issue of algorithmic discrimination, EU data protection law and in particular the GDPR can provide important complements to EU gender equality and non-discrimination law.

1.6 Conclusion

The aim of this chapter has been to clarify and define a number of basic notions related to algorithmic decision making, as well as identifying and describing characteristics and challenges of algorithms that are particularly relevant in light of the prohibition of discrimination and the right to equal treatment.

Different algorithms have been distinguished and their functions have been explained. The main distinction is one between rule-based algorithms and machine-learning/deep-learning algorithms:

- *Rule-based algorithms* are based on a fixed set of instructions, rules and variables that are programmed into a computer, and that result in highly predictable output. These algorithms can therefore be easily used for bulk decision making or ‘automated decision making’ (ADM).
- *Machine-learning and deep-learning algorithms* can be developed and ‘trained’ to analyse data in a specific way, trace correlations, find relevant patterns, make calculations, predict future behaviour, create individual profiles, and the like. Deep-learning systems can even do so independently, without being elaborately trained. Combined with statistical techniques for big data analysis, machine-learning and deep-learning algorithms can be used for *profiling* purposes as well as for *predicting* individual and group behaviour.

In practice, all types of algorithms are often connected to other algorithms and technologies (‘enabling technologies’). To the extent that this results in strongly interconnected processes that allow algorithms to perform tasks autonomously in a way that is close to what humans can do, the term artificial intelligence or AI is often used.

In relation to the application of algorithms, three different stages have been identified:

1. At the *planning stage* it is decided whether an algorithm is going to be used for decision making, what type of algorithm will be used (e.g. rule-based or machine-learning), how it is going to be used (e.g. for automated decision making or as an aid to human decision making), who is going to develop the algorithm, what datasets will be used to train the algorithm, and so on.
2. The *development stage* mainly involves the creation and training of an algorithm.
3. The *decision-making and use stage* encompasses the actual implementation of the algorithm either for direct or automated decision making, or to support human decision making. In addition, at this stage monitoring takes place to see whether the algorithm continues to generate reliable and acceptable outcomes.

At each of these stages, different individuals and organisations may play a role, and the challenges, risks and problems related to equality and non-discrimination may manifest themselves differently.

Regardless of the differences between algorithms and the different stages of algorithmic decision making, six common characteristics and challenges have been highlighted that are relevant to issues of gender equality and non-discrimination:

1. *The human factor and the stereotyping and cognitive bias challenges*
Human beings play a crucial role in all three stages of algorithmic decision making and for all types of algorithms. Humans decide on the use of a certain algorithm, they develop and train it, and they make specific use of it and monitor its performance. This human involvement may act as a check on the performance of algorithms (hence the often expressed need for a 'human in the loop'), but it may also be a risk from the perspective of equality and non-discrimination. In particular, the many flaws, biases, logical errors and fallacies inherent to human thinking and reasoning (including stereotyped thinking and thinking in terms of traditional gender roles) may influence the design and use of algorithms.
2. *The data challenge*
Algorithms can only work if they are fed with data, which are then processed to generate a certain output. Consequently, the quality, accuracy, validity and reliability of all algorithms and their output depend to a large degree on the quality, accuracy, validity and reliability of the input data. If input data is unbalanced or biased, for instance because it reflects prejudice or stereotyped thinking, the algorithm may produce output based on its detection of correlations and patterns in that data and, accordingly, may result in outcomes of algorithmic decision making that confirm or even reinforce existing patterns of social exclusion and discrimination.
3. *The correlation and proxy challenge*
Machine-learning algorithms typically identify correlations between various data points, which they use to detect patterns and make predictions. Human views on causality may differ from the correlations found by an algorithm, and some correlations found may be regarded as irrelevant or unacceptable by humans. If unquestioned and uncorrected, the emphasis on finding correlations may lead to unfoundedly discriminatory algorithmic outputs. In addition, even when protected characteristics (such as gender) are removed from the pool of available inputs, algorithms might select apparently unrelated but *de facto* correlated data points ('proxies') for prediction purposes. Indirectly, this might still lead to discriminatory outcomes.
4. *The transparency and explainability challenge*
Even for specialists, algorithmic processes may be opaque, in particular when different algorithms work together or function highly autonomously, or when their working is covered by trade secrets or intellectual property rights. This opacity and lack of information makes discrimination and bias difficult to discover, both for potential victims of discrimination and for monitoring and supervisory bodies and courts.
5. *The scale and speed challenge*
Algorithmic decisions can apply to a much wider audience than traditional human decisions, and they can be made on a much larger scale and with unprecedented speed. As a result, algorithmic discrimination can 'spread' at a wider scale and a much quicker pace than 'human' discrimination.
6. *The responsibility, liability and accountability challenge*
In all three stages of algorithmic decision making, different players are involved in setting objectives, deconstructing decision-making processes, programming and training algorithms, collecting and preparing the training data, using algorithms for decision making, monitoring and supervising their effect, etc. In addition, in many AI systems, different algorithms and enabling technologies are connected, which further increases the number of groups and individuals involved. In light of this plurality of players, it may be difficult for (potential) victims or for supervisory or monitoring bodies to know whom to hold responsible, liable and/or accountable for a discriminatory outcome.

As regards terminology, this chapter clarified that the report uses the term 'algorithmic discrimination' to refer to the types of algorithmic bias that are problematic from the point of view of EU gender equality and non-discrimination law. In addition, the terms 'equality' and 'non-discrimination' will be used to refer to the legal principles underpinning strategies against 'algorithmic discrimination' in the context of EU law instead of the ethics-oriented concept of 'algorithmic fairness'. Finally, it has been noted that the EU legislation in the data protection field may display different conceptual approaches to the issue of algorithmic discrimination, but there is also a significant overlap and correlation. Therefore, although this

report concentrates on EU gender equality and non-discrimination law, data protection legislation and approaches are discussed where relevant.

2 Challenges to the EU gender equality and non-discrimination legal framework

This second chapter discusses the risks of discrimination that arise from the increasing use of algorithms and the challenges they pose for the current EU gender equality and non-discrimination legal framework. The aim of this chapter is twofold. First, it highlights how and why algorithmic discrimination is an issue of EU gender equality and non-discrimination law. Secondly, it assesses to what extent the legal framework in place is fit for purpose, and where the gaps and challenges lie. The chapter is divided into four main sections. The first section examines the scope of EU gender and non-discrimination law in light of risks of algorithmic discrimination (section 2.1). The second section explores how algorithmic discrimination challenges the ground-based structure of EU non-discrimination protection (section 2.2). The third section investigates how algorithmic discrimination blurs the traditional doctrinal lines between direct and indirect discrimination (section 2.3), and the fourth and final section considers the issues that arise in relation to questions of evidence, responsibility and liability (section 2.4). Throughout the chapter, it is argued that although algorithms pose new discriminatory threats, most of the gaps and weaknesses that can be identified in relation to the existing legal framework are already well-known problems that have been repeatedly highlighted by legal scholars.

2.1 The scope of EU gender equality and non-discrimination law in light of the problem of algorithmic discrimination

This section discusses the issues of algorithmic discrimination that arise in relation to gender equality and non-discrimination and reviews the risks and challenges they pose in light of the current personal and material scope of the EU legal framework, offering specific examples and analyses where relevant.

The media has reported an increasing number of cases of gender discrimination performed by algorithms over recent years. There are numerous examples, many of which relate to algorithmic applications in use in the United States, such as the Apple Card algorithm, which was found to grant higher credit limits to men than to women despite the latter having higher credit scores¹⁵⁹ or Amazon's algorithmic hiring prototype, which was found to discriminate against women.¹⁶⁰ Similarly, numerous examples of algorithmic discrimination have been noted in relation to other protected grounds. A study by Obermeyer and others, for instance, shows how an algorithm used to predict patients' healthcare needs led to widespread discrimination on grounds of race.¹⁶¹ Because the algorithm used healthcare costs as a proxy for illness risks, which reflected the unequal access to healthcare services of Black and White populations in the US, Black patients were rated as less at risk than White patients for similar levels of actual illness, leading them to receive a lesser allocation of resources. Scholars have also demonstrated, for example, that the mailing service Gmail uses protected grounds such as sexual orientation or religious beliefs in order to expose users to targeted ads and recommendations.¹⁶² As will be shown in Chapter 3, many such examples of (potentially) discriminatory uses of algorithms can also be seen in the various European countries.

Despite its broad reach,¹⁶³ the EU legal framework offering protection against gender and race-based discrimination has a number of gaps and grey zones that are problematic in light of the phenomenon of

159 Gupta, AH (2019), 'Are Algorithms Sexist?', *The New York Times* (15 November), available at: www.nytimes.com/2019/11/15/us/apple-card-goldman-sachs.html.

160 Dastin, J (2018), 'Amazon scraps secret AI recruiting tool that showed bias against women' (10 October), available at: www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

161 Obermeyer Z and others, 'Dissecting racial bias in an algorithm used to manage the health of populations' (2019) 366 *Science* 447.

162 Lecuyer, M and others (2015): 'Sunlight: Fine-grained targeting detection at scale with statistical confidence' in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 554–566.

163 EU gender and race based non-discrimination law has a much broader material scope than EU non-discrimination law on grounds of age, disability, sexual orientation and religion (see subsection on material scope below).

algorithmic discrimination. The situation is even more problematic for the other protected grounds – age, disability, sexual orientation and religion – that have only limited protection under EU law.

2.1.1 *The legal framework*

EU gender equality law: equal treatment, equal pay and further guarantees

EU gender equality law includes the most comprehensive set of equality norms in EU law. In primary law, Article 2 TEU lists ‘equality between women and men’ among the values on which the EU is grounded and Article 3(3) TEU makes the promotion of gender equality one of the missions of the EU. Article 8 TFEU mainstreams the promotion of equality between men and women in all EU activities and is backed by Article 10 TFEU, which does so in relation to, among other things, the fight against sex discrimination. Article 19 TFEU lists sex among the protected grounds of discrimination under EU law. Article 153(1)(i) TFEU clarifies that in relation to social policy, the EU should ‘support and complement the activities of the Member States’ in relation to ‘equality between men and women with regard to labour market opportunities and treatment at work’. Article 157 TFEU contains provisions concerning equal pay for men and women, equal treatment at work and positive action. In addition, Article 20 of the EU Charter of Fundamental Rights proclaims the principle of equality before the law (which can be said to also encompass gender equality), Article 21 prohibits sex and other grounds of discrimination, and Article 23 of the EU Charter of Fundamental Rights proclaims the principle of equality between men and women in all areas.

These primary law provisions are given expression in a number of secondary law instruments. Most relevant to this report are Directives 2006/54/EC and 2004/113/EC. The Gender Recast Directive 2006/54/EC relates to the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation.¹⁶⁴ It pertains to gender equality in the labour market. Directive 2004/113/EC relates to the implementation of the principle of equal treatment between men and women in the access to and supply of goods and services and thus deals with gender equality in the consumption market.¹⁶⁵ Notably, it excludes the fields of media, advertising and education from its scope of application, which is problematic, as explained below.

Further instruments relating to gender equality are Directive 92/85/EEC on workplace safety and health for pregnant and breastfeeding women and women who have recently given birth,¹⁶⁶ Directive 2019/1158/EU on work-life balance,¹⁶⁷ Directive 79/7/EEC on social security¹⁶⁸ and Directive 2010/41/EU on self-employment.¹⁶⁹ The provisions in these instruments extend the protection of the principle of gender equality in matters closely related to the labour market and, as explained below, provide additional safeguards in light of existing risks of algorithmic discrimination.

164 Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast) (Gender Recast Directive) [2006] OJ L 204/23.

165 Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services [2004] OJ L 373/37.

166 Council Directive 92/85/EEC of 19 October 1992 on the introduction of measures to encourage improvements in the safety and health at work of pregnant workers and workers who have recently given birth or are breastfeeding [1992] OJ L 348/1.

167 Directive 2019/1158/EU of the European Parliament and of the Council of 20 June 2019 on work-life balance for parents and carers and repealing Council Directive 2010/18/EU (Work-Life Balance Directive) [2019] OJ L 188/79.

168 Council Directive 79/7/EEC of 19 December 1978 on the progressive implementation of the principle of equal treatment for men and women in matters of social security [1978] OJ L 6/24.

169 Directive 2010/41/EU of the European Parliament and of the Council of 7 July 2010 on the application of the principle of equal treatment between men and women engaged in an activity in a self-employed capacity and repealing Council Directive 86/613/EEC [2010] OJ L 180/1.

EU non-discrimination law: a hierarchy of protection

Beyond equality between women and men, non-discrimination is a general principle enshrined in Articles 2 and 3(3) TEU. Article 10 TFEU mainstreams the fight against discrimination in EU policies and activities. In addition to sex, racial or ethnic origin, religion or belief, disability, age and sexual orientation are all grounds protected against discrimination as set out in Article 19 TFEU. Article 21 of the EU Charter of Fundamental Rights also prohibits discrimination on grounds of racial or ethnic origin, religion or belief, disability, age and sexual orientation, along with other grounds such as colour, social origin, genetic features, language, political or any other opinion, membership of a national minority, property and birth.

In terms of secondary law, discrimination in relation to racial or ethnic origin is prohibited by Directive 2000/43/EC in employment matters, social protection, including social security and healthcare, social advantages, education and the access to and supply of goods and services.¹⁷⁰ The material scope of this directive is thus far-reaching and extends even beyond that of the gender *acquis*, since it also includes education.

The grounds of religion or belief, disability, age and sexual orientation are protected under another instrument, Directive 2000/78/EC, which, unlike the Racial Equality Directive, only in employment matters. As a result, discrimination on grounds of religion or belief, disability, age and sexual orientation is not prohibited in relation to education, social security, and access to goods and services including healthcare, housing, advertising and the media. This problem is well known among discrimination lawyers and has been referred to as constituting an undue ‘hierarchy’ of grounds in EU equality law.¹⁷¹ In 2008, the European Commission proposed a Horizontal Directive to remedy this gap, but so far, the Council has reached no agreement on this new piece of legislation.¹⁷²

*2.1.2 Equal pay, employment and self-employment**The ‘digital’ gender pay gap*

One of the areas of EU gender equality law at risk of algorithmic discrimination is equal pay. Important risks to gender equality arise when algorithms are (in)directly used to determine pay, in particular in the context of the collaborative economy and platform work. As noted above, Article 157 TFEU and Article 4 of Directive 2006/54/EC establish the principle of equal pay for women and men when performing equal work and work of equal value. However, uncertainties surround the applicability of these provisions as platform work often takes place outside the legal framework of a working contract.

Algorithms are often used by platforms and brokers in the gig economy.¹⁷³ Algorithms are also increasingly used to determine collaborative economy workers’ pay, depending on offer and demand, quality ratings, workers’ availability, etc. Notably, several factors taken into account by these algorithms can negatively affect gender equality in pay. It is common knowledge that platforms often prompt users to rate the service they receive.¹⁷⁴ Rating could be a channel through which discriminatory beliefs held by customers influence the working conditions of platform workers leading to, among other things,

170 Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin (Racial Equality Directive) [2000] OJ L 180/22.

171 See e.g. Waddington, L and Bell, M (2001), ‘More equal than others: distinguishing European Union equality directives’ 38 *Common Market Law Review* 587; Howard, E (2006), ‘The case for a considered hierarchy of grounds in EU law’ 13 *Maastricht Journal of European and Comparative Law* 445.

172 European Commission (2008), *Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation* COM(2008) 426 final, OJ C303/8 (European Union, 2008).

173 See e.g. Kullmann, M (2018), ‘Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law’.

174 In more detail, see Ranchordas, S (2019), ‘Public values, private regulators: between regulation and reputation in the sharing economy’ 13 *Law & Ethics of Human Rights*; Van Dijck, J, Poell, T and De Waal, M (2018), *The Platform Society. Public Values in a Connective World* (Oxford, Oxford University Press); Tadelis, S (2016), ‘Reputation and Feedback Systems in Online Platform Markets’ 8 *Annual Review of Economics* 321.

lower pay, less favourable working conditions or even job loss.¹⁷⁵ For instance, sex-based discriminatory stereotypes held by passengers could negatively influence the pay of Uber drivers by lowering their rating, therefore negatively influencing how many ride demands they receive through the Uber app.¹⁷⁶ Beyond discriminatory customer ratings, algorithms used to determine pay may also be able to take into account criteria such as platform workers' availability to work, their response time to customers' demands, the average amount of time they spend performing a task, etc.¹⁷⁷ Structural imbalances in gender roles and the fact that women on average spend more time on caregiving tasks than men might mean, for example, that they are less available to work flexible hours on demand because they need to juggle paid work with caregiving responsibilities. If reduced flexibility or availability is taken into account by an algorithm used to determine pay, this might influence the outcome and cause gender inequalities. Hence not only sex-based discriminatory stereotypes but also structural patterns of gender inequality might be factored into pay-computing algorithms, resulting in a 'digital' gender pay gap. This 'digital' gender pay gap has very real consequences, which have been confirmed in practice. Indeed, Barzilay and Ben-David's study on platform work shows that the average hourly income of female platform workers only amounts to two thirds of men's hourly wage.¹⁷⁸ However, the potential lack of transparency of the algorithms used to determine gig workers' pay might make it difficult to understand how sex discrimination creeps into the determination of pay and how equal work or work of equal value is understood and defined.

Algorithmic discrimination, employment and platform work

The increasing involvement of algorithms in human resources recruitment processes means that algorithmic discrimination could represent an important risk in the realm of the labour market. If uncorrected, algorithms trained on past data about promotions and recruitment will inevitably reproduce the current discriminatory status quo, thus disadvantaging legally protected groups.¹⁷⁹ While such forms of algorithmic discrimination would most likely fall under the scope of EU gender equality and non-discrimination law, algorithmic discrimination could be particularly pervasive in the context of platform work. However, the very applicability of the equal pay principle as well as further gender equality and non-discrimination guarantees linked to employment and working conditions will depend on the existence of an employment relationship between *de facto* workers and platforms or goods and services providers.¹⁸⁰

The crucial question as to the existence of that relationship has not yet been settled by the Court of Justice and uncertainties remain as regards the status of platform workers and the applicability of EU labour, social and gender equality and non-discrimination law, including the equal pay guarantees contained in Article 157 TFEU and Articles 1(b) and 4 of Directive 2006/54/EC. Several elements point in the direction of the existence of an employment relationship in some cases of platform work. However, it is important to underline that each platform operates on the basis of a different business model and therefore the extent to which it acts as an intermediary, or itself provides services, varies from case to case. In the case of Uber, for instance, the CJEU decided in *Asociación Profesional Elite Taxi* that the

175 Kullmann, M (2018), 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law', p. 8.

176 Kullmann, M (2018), 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law', pp. 8-11. See also De Groen, WP and Maselli, I (2016), *The Impact of the Collaborative Economy on the Labour Market* (Centre for European Policy Studies Special Report) 5-6; especially in the US there is much empirical evidence available on discrimination in platform ratings – see e.g. Leong, N and Belzer, A (2017), 'The new public accommodations: race discrimination in the platform economy' 105 *Georgetown Law Journal* 1271; Ge, Y and others (2016), 'Racial and gender discrimination in transportation network companies', NBER Working Paper No 22776, available at: www.nber.org/papers/w22776.

177 Kullmann, M (2018), 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law', p. 13. See more generally also Leong, N and Belzer, A (2017), 'The new public accommodations: race discrimination in the platform economy' 105 *Georgetown Law Journal* 1271; Van Dijck, J, Poell, T and De Waal, M (2018), *The Platform Society. Public Values in a Connective World* (Oxford, Oxford University Press) 78; Citron, DK and Pasquale, FA (2014), 'The scored society: due process for automated predictions' *Washington Law Review* 1, 4. Wagner and Eidenmüller even speak of 'mass production of bias' (Wagner, G and Eidenmüller, HGM (2019), 'Down by algorithms? siphoning rents, exploiting biases and shaping preferences – the dark side of personalized transactions' 86 *University of Chicago Law Review* 581, 593).

178 Barzilay, AR and Ben-David, A (2017), 'Platform Inequality: Gender in the Gig-Economy' 47 *Seton Hall Law Review* 393.

179 See e.g. Kim, PT (2017) 'Data-Driven Discrimination at Work'.

180 Caracciolo di Torella, E and McLellan, B (2018), *Gender Equality and the Collaborative Economy* (Publications Office of the European Union).

platform did not simply provide intermediary services but was to be regarded as a transport company.¹⁸¹ Although this does not directly clarify the status of Uber drivers, it is a first step towards establishing that an employment relationship exists between Uber and its drivers. Other national courts have gone further and have ruled that Uber drivers are workers, thus falling within the scope of labour law and existing equal pay guarantees.¹⁸²

Other elements of the Court's jurisprudence seem to point in the direction of an employment relationship and thus the applicability of equal pay guarantees. For instance, in *Allonby*, the CJEU clarified that the definition of a 'worker' was a matter of EU law and did not depend on national labour law definitions of workers.¹⁸³ In so ruling, the Court included 'bogus' or economically dependent but formally self-employed workers within the EU law definition of a worker, thereby rendering equal pay guarantees applicable to those workers. Further, in *Danosa*, the CJEU ruled that in order to respect the objectives of pregnancy and maternity protection set out by Directive 92/85/EEC, these guarantees needed to be extended to members of a company's board of directors, who fell under the EU law definition of workers despite the absence of a formal employment relationship at national level.¹⁸⁴ The rather purposive and autonomous definition of a 'worker' put forward by the CJEU could mean that equal pay and other gender equality guarantees, such as the protection of pregnant and breastfeeding workers, workers who have recently given birth, parenting workers and workers who assume caregiving duties contained in Directive 2006/54/EC, Directive 79/7/EEC, Directive 92/85/EEC and Directive 2019/1158/EU, might be extended to certain categories of platform workers, thus providing safeguards against direct and indirect discrimination in pay. For example, if a platform worker's availability and flexibility are criteria used by an algorithm to determine pay, the fact that a worker takes a break to breastfeed, care leave or maternity leave should not affect her level of pay or her allocation of work, such as the number of ride demands she receives. Besides, the CJEU's ruling in *Bouagnaoui* clarifies the fact that customers' (prejudiced) preferences cannot serve as an acceptable justification for discrimination in employment situations, which could serve as a basis to challenge the use of biased customer ratings in the calculation of workers' pay.¹⁸⁵ Bringing certainty to the applicability of EU social policy and gender equality law to gig economy workers is thus necessary to ensuring that the 'digital' gender pay gap and the potentially dramatic consequences of algorithmic decision making in platform work is addressed and would offer remedies to (certain categories of) platform workers.

Another regime applies to those platform workers who do not fall under the EU law definition of a worker, for example because the platform that they work for is regarded as a pure intermediary platform as opposed to a service provider. Directive 2010/41/EC on equal treatment in self-employment offers a number of gender equality guarantees to self-employed service providers. Maternity benefits, for example, must be available to self-employed platform workers.¹⁸⁶ In addition, according to Article 4 of the Directive, discrimination is prohibited in relation to 'the establishment, equipment or extension of a business or the launching or extension of any other form of self-employed activity'. However, these legal safeguards do not adequately address the problem of the 'digital' gender pay gap as equal pay guarantees will not apply to self-employed platform workers. This is an important gap in the enforceability of the principle of equal pay and a major obstacle to the EU legislature's efforts to tackle the gender pay gap.

Finally, the use of algorithmic profiling and targeting in the advertising of job-related adverts might reinforce existing patterns of discrimination in the labour market. An empirical study showed that employment ads distributed by Facebook with settings geared towards a neutral distribution ended up reaching an audience composed of 85 % women for cashier positions in supermarkets, while ads for taxi

181 Judgment of 20 December 2017, *Asociación Profesional Elite Taxi v Uber Systems Spain*, SL C-434/15 EU:C:2017:981.

182 See e.g. the UK Court of Appeal decision *Uber BV v Aslam and others* [2018] EWCA Civ 2748; the case is now pending before the UK Supreme Court. See also the French Cour d'Appel decision, *CA Paris*, 6-2, 10 January 2019.

183 Judgment of 13 January 2004, *Debra Allonby v Accrington & Rossendale College, Education Lecturing Services, trading as Protocol Professional and Secretary of State for Education and Employment* C-256/01 EU:C:2004:18.

184 Judgment of 11 November 2010, *Dita Danosa v LKB Lizings SIA* C-232/09 EU:C:2010:674.

185 Judgment of 14 March 2017, *Asma Bouagnaoui et Association de défense des droits de l'homme (ADDH) contre Micropole SA* C-188/15 EU:C:2017:204.

186 Article 8, Directive 2010/41/EU.

driver positions reached a 75 % Black audience and ads for lumberjack positions reached an audience that was 90 % male and 72 % white.¹⁸⁷ This form of stereotyping in the exposure to job adverts risks reinforcing structural inequality. While advertising is clearly excluded from the scope of the Gender Goods and Services Directive, as will be examined in the next section, advertising in relation to job positions seems to fall within the scope of access to employment. The algorithmic targeting of employment ads, if discriminatory on grounds of gender, race, age, disability, sexual orientation or religion or belief, could indeed be captured by Article 14(1)(a) of the Gender Recast Directive and Article 3(1)(a) of the Racial Equality Directive and the Framework Directive, which indicate that the prohibition of discrimination applies to ‘conditions for access to employment [...] including selection criteria and recruitment conditions’. This can be confirmed by drawing an analogy with the jurisprudence of the Court of Justice in relation to discrimination on grounds of race and sexual orientation, which is also covered by EU law on equal access to employment. In *Feryn*, *Accept* and *Associazione Avvocatura per i diritti LGBTI*, the CJEU ruled that deterring job applicants from protected groups from applying to given job positions was to be considered discrimination, even where no recruitment process was on-going.¹⁸⁸ The lack of, or reduced, advertising of, given jobs to a protected group could thus be considered discrimination if it in effect undermines the objective of EU law in terms of guaranteeing equal access to the labour market.¹⁸⁹

2.1.3 Goods and services: problematic gaps in the material scope

In the current situation, the hierarchy of protection between, on the one hand, race and gender equality, which are protected in the realm of goods and services, and, on the other hand, non-discrimination on grounds of disability, age, sexual orientation, religion or belief, which are only protected in the realm of the labour market, is highly problematic. As was discussed in Chapter 1, the availability and use of algorithms multiplies the opportunities for private companies to profile users in granular ways and target potential clients with personalised offers. Algorithmic discrimination is thus most likely to happen in the goods and services market, where users’ behaviours are analysed and responded to with differently shaped and priced offers and opportunities. As a result of this legislative gap, EU law does not protect EU citizens against algorithmic profiling and targeting in this area, which means that certain disadvantaged groups can be lawfully excluded from the access to certain goods and services. For example, one could imagine discrimination to arise in the offer of particularly vital goods and services such as housing, health, education, etc. Even though national law could prohibit such instances of discrimination, no harmonised prohibition exists at EU level. Beyond this major gap in the material scope of EU non-discrimination law, other problematic exceptions exist, in particular in relation to gender equality. These exceptions pertain to the content of media, advertising and education, which are excluded from the scope of Directive 2004/113/EC.¹⁹⁰ In light of the growing use of AI in the fields concerned, these exceptions might lead to important weaknesses in terms of the ability of EU law to redress algorithmic discrimination, as explained below.

Algorithmic discrimination in the media and advertising

As is further illustrated in Chapter 3, algorithms can easily be used in media and advertising services and gender-based algorithmic discrimination risks being pervasive in these fields. Scholarship has shown that this type of sex-based discrimination can take several forms. One such form of discrimination is

187 The authors specify that these statistics correspond to ‘the most extreme cases’ of skewed distribution. In the experiment conducted, they selected an identical audience for all three adverts. Ali, M and others (2019), ‘Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes’, arXiv preprint available at: arXiv:190402095 1.

188 Judgment of 10 July 2008, *Centrum voor gelijkheid van kansen en voor racismebestrijding v Firma Feryn NV* C-54/07 EU:C:2008:397; Judgment of 25 April 2013, *Asociația Accept v Consiliul Național pentru Combaterea Discriminării* C-81/12 EU:C:2013:275; Judgment of 23 April 2020, *NH v Associazione Avvocatura per i diritti LGBTI — Rete Lenford* C-507/18 EU:C:2020:289.

189 By analogy, Article 3(2)(b) prohibits discrimination on grounds of nationality in relation to the advertising of job positions to foreign candidates, clarifying that provisions and practices that ‘limit or restrict the advertising of vacancies in the press or through any other medium or subject it to conditions other than those applicable in respect of employers pursuing their activities in the territory of that Member State’ are unlawful.

190 Article 3(3), Directive 2004/113/EC.

harmful stereotyping.¹⁹¹ For example, Noble describes how a search engine like Google contributes to forging and maintaining harmful sexist stereotypes by returning mostly female pictures on an image search like 'nurse' and mostly male pictures on a search like 'doctor'.¹⁹² She demonstrates how sex stereotyping by these search algorithms is often intersectional, taking the example of an image search for 'professional hairstyles for work' returning mostly pictures of White women's hairstyles, while an image search for 'unprofessional hairstyle for work' returned mostly images of Black women's hairstyles. Given the dominant position of Google among search engines and the crucial functions it fulfils in terms of learning, knowledge discovery and access to information, such misrepresentation issues and stereotyping are problematic and play a role in reinforcing and spreading discriminatory beliefs. While Google's defence has been that its search engine merely reflects users' beliefs, this explanation does not seem satisfactory in light of the company's considerable influence on the production, creation and presentation of knowledge, beliefs and opinions. Another study by Kay et al. shows how the occupational representation of women and men in online search results reflects gender segregation in the labour market and reinforces gender stereotypes.¹⁹³

Since the media are excluded from Directive 2004/113/EC, these types of representational intersectional and sex discrimination are out of reach of EU gender equality law. Nevertheless, this problem could be tackled at national level in Member States that have not implemented these exceptions and whose gender equality law goes further than that of EU law.¹⁹⁴

A similar scope issue arises in relation to sex discrimination in online advertising, which is excluded from the scope of Directive 2004/113/EC. Algorithms play an increasing role in the targeted distribution of ads to users of online platforms such as Facebook, Twitter or Instagram and search engines like Google. Research on the algorithmic optimisation of advert distribution for example shows how gender, race and other stereotypes shape the distribution of personalised ads by Facebook to its users, even when the target audience is set as neutral in relation to these characteristics by advertisers.¹⁹⁵ Another study demonstrates how gender also plays an indirect role in determining which end users will be exposed to which online ads. Because women are a more valuable target than men for marketing purposes, it is more expensive to advertise to them and they are shown higher bid ads, which results in skewed exposure to advertisements even when ad delivery settings are set as neutral.¹⁹⁶ An empirical experiment conducted with an advert for information on STEM careers showed a higher display of the ad to men than to women, despite neutral targeting settings, which risks reinforcing existing patterns of gender-based labour segregation as well as gender stereotypes.¹⁹⁷ Commentators highlight that solutions to this problem of discriminatory advertising might be hard to find in the current circumstances given that platforms' ad targeting policies do not let advertisers choose gender as a distribution criterion for job-

191 Stereotyping can be harmful for various reasons, for instance when undermining dignity, preventing access to certain goods, services or social recognition, maintaining gender segregation by prescribing certain roles and maintaining given expectations, etc. See e.g. Timmer, A (2016) *Gender Stereotyping in the Case Law of the EU Court of Justice*, pp. 38-9.

192 Noble, SU (2018), *Algorithms of oppression: how search engines reinforce racism* (New York University Press) 82-3.

193 See Kay, M; Matuszek, C; Munson, S (2015) 'Unequal representation and gender stereotypes in image search results for occupations' in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3819-3828, published by ACM.

194 For more information, see Burri, S; Senden, L and Timmer, A (2020) *A comparative analysis of gender equality law in Europe 2019* (Luxembourg: Publications Office of the European Union).

195 Ali, M and others (2019), 'Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes', arXiv preprint available at: arXiv:190402095 1.

196 The cost of advertising to given population groups is calculated, for instance, as a function of their likeliness to click on the advert and/or purchase the goods or service advertised. Women are a more valuable target because they make more purchasing decisions regarding household goods than men do. Adverts are distributed based on bids, i.e. the amount of money an advertiser is willing to pay for each click, follow-up purchase, etc. by an online user. Valuable advertising groups are thus shown higher bid ads. Lambrecht, A and Tucker, C (2019), 'Algorithmic Bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads' 65 *Management Science*, pp. 2966-7.

197 Lambrecht, A and Tucker, C (2019), 'Algorithmic Bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads'.

related ads, even where the aim is to obtain a more equal distribution of ads among different genders.¹⁹⁸ Some studies have suggested that platforms could solve this problem by offering advertisers the option to choose an equalised distribution of ads across different population groups.¹⁹⁹

When such ads concern goods and services, it could be argued that they are excluded from the scope of EU gender equality law given the provisions in Article 3(3) of Directive 2004/113/EC. At the same time, it could be argued that discrimination arises because the lack of, or reduced, advertising of given goods and services (e.g. health or housing) to certain protected groups (e.g. women) might hinder their access to these goods and services. Supporting this argument is the fact that only the ‘content of [...] advertising’ is excluded from the scope of the Directive, and not its distribution.²⁰⁰ This would fit with the principle of effectiveness of EU law, since the discriminatory advertising of goods and services, for housing or health-related services for instance, could undermine the objective of banning such discrimination in the consumption of these goods and services. Another risk would be the reinforcing of existing harmful stereotypes and gender segregation if advertising patterns contribute to maintaining discriminatory gender roles. One could imagine, for example, that the over-exposure of women to ads for goods and services linked to caregiving and the home could contribute to reinforcing existing role-typing and prescriptive stereotypes about the role of women as homemakers and caregivers.²⁰¹ However, in the absence of case law, uncertainties exist regarding whether or not this would fall within the scope of EU law.

Beyond the specific weaknesses pointed out above in relation to gender equality and the total absence of coverage in relation to discrimination on grounds of age, disability, sexual orientation and religion in this field, algorithmic discrimination in relation to race or ethnic origin raises some questions. As explained above, harmful stereotyping and prejudices could pervade algorithms used to determine the distribution of ads and ultimately the access to goods and services.²⁰² For instance, research has shown that housing ads distributed by platforms using algorithms without inputting race as a target criterion could discriminate against ethnic groups.²⁰³ In addition, attention has been drawn to how indexing algorithms used by online search engines can perpetuate racism through returning results that reflect stereotypes.²⁰⁴ Although race equality is covered by EU law in the access to and supply of goods and services, no mention is made of media and advertising. Since there is no explicit exception as in Directive 2004/113/EC, it would be logical to assume that these fields are covered by the Racial Equality Directive 2000/43/EC. However, certainty over the inclusion of these fields in the material scope of the Directive would be welcome given the risks of algorithmic discrimination arising in media and advertising.

198 See Lambrecht, A and Tucker, C (2019), ‘Algorithmic Bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads’, p. 2278: ‘Facebook did not approve these ads as they do not allow advertisers to exclude users of either gender when running an employment-related ad’.

199 Lambrecht, A and Tucker, C (2019), ‘Algorithmic Bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads’, p. 2278.

200 This links up to broader questions of discrimination through ‘gender versioning’, a commercial technique by which similar goods and services consumed by both men and women are artificially differentiated on grounds of gender and thereby sold to women at a higher price. This is often the case for cosmetics and beauty-related goods and services. See e.g. De Blasio, B and Menin, J (2015), *From cradle to cane: the cost of being a female consumer. A study of gender pricing in New York City* (New York Department of Consumer Affairs). Through gender versioning, products could illegitimately fall within the exception set out by Article 4(5), which states that the Directive ‘shall not preclude differences in treatment, if the provision of the goods and services exclusively or primarily to members of one sex is justified by a legitimate aim and the means of achieving that aim are appropriate and necessary’.

201 Researchers have proven that such stereotypes are deeply engrained in language and that language-processing algorithms learned and reproduced them automatically. See e.g. Bolukbasi, T and others (2016), ‘Man is to computer programmer as woman is to homemaker? Debiasing word embeddings’ *Advances in Neural Information Processing Systems* 4356.

202 See Sweeney, L (2013) ‘Discrimination in Online Ad Delivery’ in *Communication of the ACM* 56(5), 44.

203 Angwin, J and Parris Jr, T (2016), ‘Facebook lets advertisers exclude users by race’, *ProPublica* available at: <http://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race> and Gabbatt, A (2019), ‘Facebook charged with housing discrimination in targeted ads’, *The Guardian* (28 March) available at: www.theguardian.com/technology/2019/mar/28/facebook-ads-housing-discrimination-charges-us-government-hud.

204 See Noble, SU (2018), *Algorithms of oppression: how search engines reinforce racism* (New York University Press).

Algorithmic discrimination in education

In the field of education, too, algorithms are increasingly used to assist decision making, as the examples discussed in Chapter 3 may illustrate. For example, France has invested in an algorithm to support decision making in relation to the allocation of places in higher education institutions to incoming students (Parcoursup). Concerns were expressed by candidates regarding potential risks of discrimination and the lack of transparency in the decision-making process. These concerns regarding the use of the Parcoursup algorithm related to the consequences of the use of income and residency data about candidates in the allocation decisions.²⁰⁵ The generalisation of such algorithmically assisted recruitment processes in the field of education could lead to algorithmic discrimination if not kept in check. However, at present, only the grounds of race and ethnic origin are protected against discrimination in education in the EU. Education is indeed an exception to the material scope of EU gender equality law. This further gap could prove problematic from the perspective of redress in Member States that have implemented that exception.

Furthermore, the exclusion of education from the scope of Directive 2004/113/EC is problematic in light of the under-representation of women in STEM fields and curricula related to IT and software development, which is specifically discussed in section 3.2.6. The lack of representation of women in these fields of education maintains gender segregation at a later stage in the labour market and leads to a lack of diversity in software developers and programmers, which in turn means that algorithms fail to reflect a variety of gender perspectives. The lack of EU legal guarantees against discrimination on grounds of age, disability, sexual orientation and religion is problematic for the same reasons. This lack of voice of women and minority groups in algorithmic design has clear repercussions in terms of biased algorithmic design leading to discrimination. For example, the case of Dr Selby illustrates how harmful gender stereotypes crept into the design of a piece of commercial software giving automated access to changing rooms in a fitness studio.²⁰⁶ Because Dr Selby's title was 'Dr' and not 'Ms', she had been classified as a man and could not enter the women's changing rooms. This type of mistake in the design of algorithms could be corrected through increasing the diversity of the workforce, and thus the representation of various minority perspectives, in relevant areas of the labour market. This shows that the sources of algorithmic discrimination lie as much in the functioning of algorithms as in their human design.

Enforcement issues: algorithmic pricing and discrimination

The use of algorithms in the market for goods and services, also termed algorithmic pricing, can lead to discrimination if it takes protected characteristics into account in the determination of goods and services' prices, for instance when users' data are used to personalise prices.²⁰⁷ While this should theoretically be captured by the Gender Goods and Services Directive and the Racial Equality Directive, as confirmed by the CJEU in relation to the use of gender as an actuarial factor in insurance policy pricing in the *Test-Achats* case, doubts nonetheless arise in three regards. First, the tailoring of goods and services and the personalisation of pricing could cast doubt on whether a given good or service is indeed 'available to the public', therefore potentially rendering the relevant legal guarantees in Directive 2004/113/EC and Directive 2000/43/EC inapplicable.²⁰⁸ Secondly, in relation to gender in particular, gender versioning could be used to let given goods and services fall within the provisions of Article 4(5) of Directive 2004/113/EC,

205 See Défenseur des Droits (2019) 'Parcoursup: le Défenseur des droits, dans deux décisions, recommande plus de transparence, de mobilité et de mixité et des mesures d'accompagnement adaptées pour les personnes handicapées' (Parcoursup: in two decisions, the Defender of Rights recommends more transparency, mobility and diversity as well as appropriate support measures for people with disabilities), available at: <https://www.defenseurdesdroits.fr/fr/communiqu%C3%A9-de-presse/2019/01/parcoursup-le-defenseur-des-droits-dans-deux-decisions-recommande-plus>.

206 Turk, V (2015), 'When Algorithms Are Sexist', *Vice* (20 March) available at: www.vice.com/en_us/article/ezvkee/when-algorithms-are-sexist.

207 See e.g. Zuiderveen Borgesius, F and Poort, J (2017), 'Online Price Discrimination and EU Data Privacy Law' *Journal of Consumer Policy* 363.

208 See Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law' 55 *Common Market Law Review* 1143.

whereby the Directive ‘shall not preclude differences in treatment, if the provision of the goods and services exclusively or primarily to members of one sex is justified by a legitimate aim and the means of achieving that aim are appropriate and necessary’. Finally, it is well known that the enforcement of the prohibition of gender-based price discrimination at national level is difficult to guarantee. Several experts point to deficiencies in this regard and this type of discrimination is sometimes regarded as too trivial for action to be taken.²⁰⁹ Often-quoted examples include the differential pricing of haircuts or entry tickets to clubs for men and women.²¹⁰ Such enforcement difficulties might increase with dynamic pricing algorithms that constantly adapt the price of given goods and services.

All in all, this section has shown that a number of problematic gaps exist in the material scope of EU gender equality and non-discrimination law, which question its capacity to adequately capture, address and redress algorithmic discrimination. The next section explores whether the personal scope – in other words, the protected grounds – is fit to address problems of algorithmic discrimination.

2.2 Protected grounds and algorithmic discrimination

Beyond the gaps in the material scope of EU gender equality and non-discrimination law described above, uncertainties in the personal scope create further weaknesses in light of the problem of algorithmic discrimination. As explained above, the EU gender equality and non-discrimination directives are characterised by their closed lists of grounds: discrimination is prohibited only if it can be shown to be based on sex, race/ethnic origin, religion/belief, disability, sexual orientation and age or to disproportionately disadvantage a person or group identified by one of the listed characteristics. The forms of discrimination arising from the use of algorithms set some particular challenges in relation to the protected grounds that define the personal scope of EU equality law.

2.2.1 Algorithmic gender-based classification

To start with, the categorisation of users performed by profiling algorithms might raise issues of discrimination in and of itself. Because the operation of algorithms relies on classification and categorisation, risks of discrimination exist in relation to the inclusion and exclusion of individuals in and from given groups. Gender-based algorithmic classification could for instance lead to excluding gender non-conforming, trans and intersex individuals from access to certain goods, services and jobs or forcing them into categories with which they do not identify. The same is true for biometric border control technologies, which enforce an artificial gender binary where reality shows the existence of a gender continuum.²¹¹ Algorithmic gender-based classification could thus lead to grave situations of discrimination. This poses questions regarding the personal scope of EU gender equality law. On the one hand, the Court of Justice has long recognised that EU gender equality law applies to situations of discrimination arising from ‘gender reassignment’.²¹² Algorithmic discrimination in this regard would thus be captured by the legal framework in place. On the other hand, it is uncertain whether EU gender equality law protects intersex and gender non-conforming persons from discrimination. Indeed, it does not explicitly include gender identity, gender expression or sex characteristics in its personal scope. In addition, in protecting the

209 See Burri, S and McColgan, A (2008), *Sex-segregated Services* (European Commission) and Burri, S and McColgan, A (2009), *Sex Discrimination in the Access to and Supply of Goods and Services and the Transposition of Directive 2004/113/EC* (European Commission). At the same time a recent campaign in the Netherlands has drawn attention to the issue of gender pricing in hairdressers’ services and led the Institute for Human Rights to find discrimination in two cases, see College voor de Rechten van de Mens (2020), ‘College oordeelt over verschil in kapperstarieven voor mannen en vrouwen’ (Human Rights Institute judges on difference in hairdresser’s tariffs for men and women) (16 March 2020), available at: <https://mensenrechten.nl/nl/nieuws/college-oordeelt-over-verschil-kapperstarieven-voor-mannen-en-vrouwen>.

210 See e.g. European Commission (2015) ‘Report on the application of Council Directive 2004/113/EC implementing the principle of equal treatment between men and women in the access to and supply of goods and services’ COM(2015) 190 final (Brussels: European Union), 7.

211 See the GIRARE project by Van den Brink, M and Quinan, C, available at: www.uu.nl/en/research/netherlands-institute-of-human-rights-sim/research/girare.

212 See Judgment of 30 April 1996, *P v S and Cornwall County Council* C-13/94 EU:C:1996:170, [20].

equality rights of trans persons, the Court of Justice has arguably reinforced the gender binary instead of making space for non-binary identities.²¹³ Hence, there is no certainty regarding the protection offered by EU law to intersex and gender non-conforming persons in case of gender-based discrimination arising from algorithmic decision-making procedures. The same can be said where algorithmic discrimination is intersectional and involves gender. Despite the existence of recitals mentioning ‘multiple discrimination against women’ in Directives 2000/43/EC and 2000/78/EC, the CJEU failed to explicitly recognise the existence of intersectional discrimination in *Parris*.²¹⁴ Neither Directive 2004/113/EC nor Directive 2006/54/EC mentions intersectional or multiple discrimination. While it is likely that algorithmic profiling and targeting leads to an increase in cases of intersectional discrimination,²¹⁵ doubts concerning the personal scope of EU gender equality law make it difficult to predict whether such situations could be adequately addressed by EU law.²¹⁶ In light of the rise of algorithmic profiling technologies, it seems important to adopt an expansive reading of the personal scope of EU gender equality and non-discrimination law so as to protect individuals and groups whose identities are intersectional and non-binary.

2.2.2 Correlations and proxies

In addition to the difficulties linked to algorithmic classification, section 1.4.3 has highlighted what we called the correlation and proxy challenge. This challenge describes the fact that even if algorithms can be trained to reject protected grounds as irrelevant for finding patterns and correlations, they can still detect so-called proxy variables that are very closely related to the protected grounds.²¹⁷ In fact, it would be quite rare for an algorithm to directly and openly discriminate only or decisively on the basis of a protected ground, since it will usually base its output on a multitude of different factors and variables that are all statistically correlated.²¹⁸ As a consequence, the basis of a decision will be granular and diverse and thus might be difficult to relate to a particular protected ground.²¹⁹ Moreover, due to what we call the transparency challenge, it may be difficult to identify exactly which variables explain a certain algorithmic output, making it even harder to detect and isolate the ‘actual’ ground of discrimination (as well as making it easier to ‘mask’ it).²²⁰

Beyond the identification of protected grounds in algorithmic decision making, the correlation and proxy challenge is relevant in relation to the definition of the protected grounds.²²¹ It raises the question of how narrowly or widely the protected grounds should be defined. If protected grounds are given an extensive interpretation, relevant proxies could also be covered by their meaning. The classic example is pregnancy-related discrimination, which is so clearly and closely related to sex-based discrimination that ‘pregnancy’ is generally regarded as a proxy for ‘being a woman’. In fact, the Court of Justice of the EU has treated

213 See Van den Brink, M and Dunne, P (2018), *Trans and intersex equality rights in Europe – a comparative analysis* (Universiteit Utrecht) 53-4.

214 Judgment of 24 November 2016, *David L. Parris v. Trinity College Dublin and Others*, C-443/15, EU:C:2016:897.

215 See also section 2.2.

216 In addition, in cases like *Achbita* and *Bougnaoui*, the CJEU has shown no sensitivity for the issue of intersectionality. See Judgment of 14 March 2017, *Samira Achbita and Centrum voor gelijkheid van kansen en voor racismebestrijding v G4S Secure Solutions NV* C-157/15 EU:C:2017:203 and Judgment of 14 March 2017, *Asma Bougnaoui et Association de défense des droits de l’homme (ADDH) contre Micropole SA* C-188/15 EU:C:2017:204.

217 See Kim, P (2017), ‘Data-Driven Discrimination at Work’ 58 *William and Mary Law Review* 857, 880.

218 Gillis, TB and Spiess, JL (2019), ‘Big Data and Discrimination’ 86 *University of Chicago Law Review* 459, 469; Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’ 55 *Common Market Law Review* 1143, 1151; compare also Crawford, K and Whittaker, M (2016), *The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*, 7, who speak of the risk of discrimination becoming much more ‘fine-grained’ as a result of the ability of algorithms to discriminate on many more factors than just the protected grounds.

219 Barocas, S and Selbst, A (2016), ‘Big Data’s Disparate Impact’ 104 *California Law Review* 761, 691.

220 See above, section 1.4.4; see also e.g. Gillis, TB and Spiess, JL (2019), ‘Big Data and Discrimination’ 86 *University of Chicago Law Review* 459, 479; Barocas, S and Selbst, A (2016), ‘Big Data’s Disparate Impact’ 104 *California Law Review* 761, 693.

221 Compare e.g. Kullmann, M (2019), ‘Discriminating job applicants through algorithmic decision-making’ *Ars Aequi* 51; Wachter, S (2020), ‘Affinity profiling and discrimination by association in online behavioural advertising’ 35 *Berkeley Technology Law Journal* 7; Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’ 55 *Common Market Law Review* 1143, 1149.

pregnancy-related discrimination as direct sex discrimination.²²² Another classic example is holding a foreign passport, which is a clear proxy for 'being of a different nationality'. Usually there is an almost 100 % overlap here between the 'actual' protected ground and its proxies, meaning that the use of the proxy covers almost exactly the same group of persons as using the actual ground would do. Similarly, when there is a close connection between individual preferences and affinities, and protected grounds, belonging to a group with a certain 'affinity' (e.g. having an interest in particular religious matters) might be nearly the same as belonging to a group characterised by a particular personal trait (e.g. adhering to a certain religion).²²³

The challenge set by algorithmic discrimination is to discover which (combinations of) proxy variables and affinities actually can be seen to have such a large degree of overlap with the corresponding protected ground that they actually can be seen as the same thing.²²⁴ In line with this, the question may arise if really a (nearly) 100 % overlap is required, or if it would be enough to show statistically that a certain variable (or a certain combination of variables) has a 90 % or 80 % overlap with a given protected ground. The question of defining the protected grounds therefore also comes close to identifying when a case of unequal treatment can still be typified as direct discrimination on a certain ground, or when it should be termed indirect discrimination, or, perhaps, as discrimination by association with a group that is characterised by a protected ground.²²⁵ Section 2.3 further examines how the use of proxies and correlations affects the continued relevance of the traditional concepts of direct and indirect discrimination.

2.2.3 New forms and grounds of discrimination

Another question that might arise in relation to algorithmic discrimination is whether the current lists of grounds contained in the EU legislation are still up to date. Some scholars have shown that algorithmic decision making has revealed patterns of structural discrimination that are based on existing inequalities related to characteristics such as socioeconomic status, education, health status and income, which is particularly problematic when these characteristics are not protected under non-discrimination law.²²⁶ In addition, it has been shown that algorithms may easily discriminate on the basis of seemingly irrelevant characteristics, such as being a dog-owner, simply because there is a correlation between owning a dog and being open to certain forms of advertising.²²⁷ This also prompts the question of the performativity of algorithmic discrimination, i.e., the question of whether the increasing use of algorithms may create *new* forms and types of discrimination.²²⁸ If this is the case, possibly new grounds of discrimination should be added to cover such new types of discrimination, or the closed system of grounds should be opened up in order to allow for such new forms.

222 See the landmark case: Judgment of 8 November 1990, *Elisabeth Johanna Pacifica Dekker v Stichting Vormingscentrum voor Jong Volwassenen (VJV-Centrum) Plus* C-177/88, EU:C:1990:383.

223 Wachter, S (2020), 'Affinity profiling and discrimination by association in online behavioural advertising' 35 *Berkeley Technology Law Journal* 7; compare also Katyal, SK (2019), 'Private accountability in the age of artificial intelligence' 66 *UCLA Law Review* 54, 66.

224 Compare Wachter, S (2020), 'Affinity profiling and discrimination by association in online behavioural advertising'; 30; Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law', 1143, 1152.

225 Wachter, S (2020), 'Affinity profiling and discrimination by association in online behavioural advertising'; Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law', 1152.

226 See e.g. Eubanks, V (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St Martin's Press); Wachter, S (2020), 'Affinity profiling and discrimination by association in online behavioural advertising', 54.

227 Wachter, S (2020), 'Affinity profiling and discrimination by association in online behavioural advertising', 56-7.

228 Xenidis, R (2020), 'Tuning EU equality law to algorithmic discrimination: Three pathways to resilience' 27(6) *Maastricht Journal of European and Comparative Law* 736.

2.2.4 The CJEU's interpretation of the grounds listed in Article 21 of the Charter of Fundamental Rights

Algorithmic discrimination thus challenges the current boundaries of EU non-discrimination law. From this perspective, the role of Article 21 of the Charter of Fundamental Rights (the Charter) in the EU equality law framework deserves some closer attention. Article 21 establishes a non-exhaustive and open-ended list of discrimination grounds by prohibiting discrimination 'based on any ground such as' the characteristics listed.²²⁹ Although, since 2009, the Charter has had the same value as the Treaties and despite the proclamation of a general principle of non-discrimination with horizontal direct effect in *Mangold* and the clarification that Article 21 of the Charter has direct horizontal effect in *Egenberger*, the Court of Justice clarified early on, in *Chacon Navas*, *Coleman* and *Kaltoft*, that only those grounds that find expression in secondary law can be held to be protected by the Directives.²³⁰ Thus, in *Coleman* and *Kaltoft*, the *Chacon Navas* line of reasoning was upheld despite the binding status acquired by the Charter inbetween the two judgments and in spite of its non-exhaustive list of protected grounds. In substance, the Court stated that 'the scope of Directive 2000/78 should not be extended by analogy beyond the discrimination based on the grounds listed exhaustively in Article 1 thereof'.²³¹ As a result, in *Kaltoft*, the CJEU held that 'obesity [could not] as such be regarded as a ground in addition to those in relation to which Directive 2000/78 prohibits discrimination' and could only be protected insofar as it could be understood through the prism of a ground already protected under EU law, in this case disability.²³² As a result, the Court curtailed the potential of Article 21 as a basis for introducing more flexibility in the personal scope of EU equality law. The exhaustive nature of the list of protected grounds in EU law and the limits put by the CJEU to their expansive interpretation raise problems in relation to proxy discrimination, an issue that is particularly acute in respect of algorithms, as explained above. Arguably, however, a broad interpretation of Article 21 of the Charter could help better capture the specific types of discrimination arising from the use of algorithms, as will be further examined in section 4.4.2.1.

2.2.5 Algorithmic granularity and intersectionality

The high level of differentiation that is enabled by algorithmic pattern analysis and profiling breathes new air into the debate on intersectional and multiple discrimination. In many cases the output of an algorithm and the decision based on it will not be based *only* on sex or *only* on ethnic origin, but on a combination of characteristics and behaviour that is unique to a particular person, or perhaps to a small group of persons. Even more than in the past, the application of an algorithm might lead to a decision that is based on a combination of several characteristics, for instance a person being a female wheelchair user belonging to an ethnic minority, or being an elderly gay man, without it being possible to identify which of these characteristics was most important in making that decision. Even though intersectional discrimination has long been recognised as a legal problem,²³³ EU law still grapples with the question of its redress. In its decision in *Parris*, the CJEU on the one hand recognised the existence of multiple discrimination, stating that 'discrimination may indeed be based on several of the grounds' protected under EU law, but on the other hand it rejected a finding of intersectional discrimination, declaring that 'no new category of discrimination resulting from the combination of more than one of those grounds [...] may be found to

229 See section 2.1.1 for a list.

230 Judgment of 22 November 2005, *Werner Mangold v Rüdiger Helm* C-144/04 EU:C:2005:709; Judgment of 17 April 2018, *Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung* eV C-414/16 EU:C:2018:257; Judgment of 11 July 2006, *Sonia Chacón Navas v Eurest Colectividades SA* C-13/05 EU:C:2006:456; Judgment of 18 December 2014, *Fag og Arbejde (FOA) v Kommunernes Landsforening (KL)* C-354/13 EU:C:2014:2463.

231 Judgment of 18 December 2014, *Fag og Arbejde (FOA) v Kommunernes Landsforening (KL)* C-354/13 EU:C:2014:2463 [36]; Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law* C-303/06 EU:C:2008:415 [46] and Judgment of 11 July 2006, *Sonia Chacón Navas v Eurest Colectividades SA* C-13/05 EU:C:2006:456 [56].

232 Judgment of 18 December 2014, *Fag og Arbejde (FOA) v Kommunernes Landsforening (KL)* C-354/13 EU:C:2014:2463 [37].

233 See e.g. Crenshaw, K (1989) 'Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics', *University of Chicago Legal Forum* 139 and in the ambit of EU law, e.g. Schiek, D (2005) 'Broadening the scope and the norms of EU gender equality law: towards a multidimensional conception of equality law', 12 *Maastricht Journal of European and Comparative Law* 427.

exist where discrimination on the basis of those grounds taken in isolation has not been established'.²³⁴ Inherent in the notion of intersectional discrimination is the fact that the discriminatory harm might not exist in relation to a sole protected ground taken in isolation, but rather only in relation to a combination of protected grounds. In *Parris*, for example, intersectional discrimination arose from the fact that the applicant, who was already over 60 when civil partnership was legalised for same-sex couples in Ireland, could not register for a survivor pension scheme because a rule meant to prevent abuse set 60 as the age limit for entering the scheme. The resulting discrimination was the product of the intersection between the age limit rule and structural discrimination on grounds of sexual orientation in relation to the regulation of social relationships. The outcome of such an intersection was a particular disadvantage pertaining to a particular group of population, namely same-sex couples older than 60. The reasoning of the Court in *Parris* however ignored the existence of such intersectional discrimination by requiring evidence of discrimination based on each protected ground involved separately.²³⁵ This evidentiary requirement arises from the comparator-based test performed by the Court in non-discrimination cases: in order to assess the existence of a difference in treatment or a disadvantage, a comparison is established between the group that shares a given protected characteristic and the group that does not. This comparator-based test has been criticised for its unidimensional or 'single-axis' nature, that is, its focus on a single ground at a time.²³⁶ The lack of redress for intersectional discrimination in EU law – despite the recognition of the issue of 'multiple discrimination' in Directives 2000/78/EC and 2000/43/EC – is particularly problematic in light of the increasing risks of intersectional discrimination linked to the granular profiling abilities of algorithms. That said, the judgment in *Parris* contrasts with other decisions of the CJEU in which, although not recognising intersectional discrimination explicitly, the Court acknowledges the disadvantageous effects of a combination of different grounds of discrimination.²³⁷ Section 4.4.2 shows how these precedents offer a legal basis to tackle this issue, and proposes potential legal solutions in relation to the architecture of EU non-discrimination legal provisions.

2.2.6 The dynamic nature of algorithmic categorisations

Finally, a remaining challenge that arises from algorithmic discrimination in relation to the grounds protected under EU law is the dynamic nature of algorithmic categorisations. In machine learning, in particular, algorithms evolve over time as they 'learn' and so does their pattern-recognition function. Hence, the relationship between algorithmic output and protected grounds and their degree of overlap might change over time as a given algorithm evolves. As Kullmann puts it, taking the example of gender, 'depending on the algorithm, the categorical membership of one or more constructions of gender can be dynamic', since different algorithms may identify and highlight different individual characteristics and preferences.²³⁸ Again, this raises the question of whether legislation that strongly differentiates between the different grounds in terms of material scope and exemptions is desirable.

234 Judgment of 24 November 2016, *David L. Parris v. Trinity College Dublin and Others*, C-443/15, EU:C:2016:897, [80].

235 See Xenidis, R (2018) 'Multiple discrimination in EU anti-discrimination law: towards redressing complex inequality?' in Belavusau, U and Henrard, K (eds), *EU anti-discrimination law beyond gender* (Hart Publishing 2018), p. 59 and p. 72.

236 See e.g. Goldberg, S (2011), 'Discrimination by Comparison' 120 *Yale Law Journal* 728; Atrey, S (2018), 'Comparison in intersectional discrimination', *Legal Studies*, 38(3), 379-395 and Jonker, M (2015) 'Comparators in multiple discrimination cases: a real problem or just a theory?' in van den Brink, M, Burri, S and Goldschmidt, J (eds), *Equality and human rights: nothing but trouble? Liber amicorum Titia Loenen*, vol 38 (Netherlands Institute of Human Rights (SIM)).

237 Judgment of 6 December 2012, *Johann Odar v Baxter Deutschland GmbH*, C-152/11, EU:C:2012:772, [69] and Judgment of 19 September 2018, *Surjit Singh Bedi v Bundesrepublik Deutschland and Bundesrepublik Deutschland in Prozessstandschaft für das Vereinigte Königreich von Großbritannien und Nordirland*, C-312/17, EU:C:2018:734, [75]. See Xenidis, R (2020), 'Tuning EU equality law to algorithmic discrimination: Three pathways to resilience', p. 744 and for an analysis of the 'intra-categorical approach' to intersectional discrimination adopted by the Court, see Xenidis, R (2020) *Beyond the 'Master's Tools': Putting Intersectionality to Work in European Non-Discrimination Law. A Study of the European Union and the Council of Europe Non-Discrimination Law Regimes* (Florence: European University Institute) and Xenidis, R (2018) 'Multiple discrimination in EU anti-discrimination law: towards redressing complex inequality?' in Belavusau, U and Henrard, K (eds), *EU anti-discrimination law beyond gender* (Hart Publishing 2018).

238 Kullmann, M (2018), 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law'.

2.3 The types of discrimination defined in EU law²³⁹

Algorithmic discrimination does not neatly fit the central concepts of EU gender equality and non-discrimination law. In particular, as was also briefly mentioned in section 2.2, the conceptual grasp of the notion of direct discrimination is decreasing in the face of the specific operation of algorithms and notably the phenomenon of proxy discrimination. In turn, the notion of indirect discrimination provides a better conceptual fit but opens a much wider pool of available justifications, thus casting doubts on the effectiveness of EU law in redressing algorithmic discrimination. This section evaluates the adequacy of these concepts as they have been applied by the CJEU for the purpose of redressing algorithmic discrimination. We argue that the actual boundaries between the two concepts become blurred in the context of algorithmic discrimination.²⁴⁰

2.3.1 Direct discrimination: an uneasy fit with algorithmic discrimination

EU law defines direct discrimination as a situation in which ‘one person is treated less favourably than another is, has been or would be treated in a comparable situation’ on the basis of one of the protected grounds defined in the relevant directives.²⁴¹ Direct discrimination focuses on ‘unfavourable treatment’ or ‘differential treatment’ and captures situations in which a decision is made taking into consideration a protected ground, to the disadvantage of the person or group of persons related to that protected ground. This concept adequately captures a wide range of situations of discrimination when performed by humans and presents a number of strengths, but also weaknesses, which are relevant in the context of algorithmic discrimination.

First, the notion of intent or intentionality is irrelevant to direct discrimination in EU law. Hence, proving a case of direct discrimination requires neither showing that the perpetrator was conscious of the discrimination nor showing that he or she intended to discriminate.²⁴² The absence of a requirement of intent means that the concept of direct discrimination potentially covers situations where the developers of an algorithm did not intend to build a discriminatory model but designed it in a way that allows the algorithm to treat individuals and groups sharing certain protected categories in a less favourable way than others. This could be the result, for example, of developers’ own biases or the use of already biased data to train a given algorithm. In that case, it would suffice to prove that a protected category plays a role in the algorithmic treatment and results in a disadvantage to the protected group in order to demonstrate the existence of direct discrimination.²⁴³ The developers’ knowledge of the discriminatory nature of the algorithm would be irrelevant.

A second strength of the concept of direct discrimination is that it extends to situations where a person is treated unfavourably because he or she is associated with a protected group, without sharing the protected characteristic himself or herself. This approach has been termed discrimination by association and has been developed by the CJEU in *Coleman*, where an employee was treated less favourably by her employer because she had to care for her disabled child.²⁴⁴ Although she did not live with disabilities herself, the Court recognised that Ms Coleman had been harassed and directly discriminated against because of her relationship or association with her disabled child.²⁴⁵ In particular, it reasoned that ‘it

239 See e.g., Balkin, JM (2017), ‘2016 Sidley Austin distinguished lecture on big data law and policy: the three laws of robotics in the age of big data’ *Ohio State Law Journal* 1217 and Pasquale, F (2017), ‘Toward a fourth law of robotics: Preserving attribution, responsibility, and explainability in an algorithmic society’ *78 Ohio State Law Journal* 1243.

240 See Xenidis, R (2020), ‘Tuning EU equality law to algorithmic discrimination: Three pathways to resilience’ 27(6) *Maastricht Journal of European and Comparative Law* 736 and Xenidis, R, ‘Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law’ (on file with the author).

241 Article 2(2)(a) Directive 2000/43/EC; Article 2(2)(a) Directive 2000/78/EC; Article 2(a) Directive 2004/113/EC and Article 2(1)(a) Directive 2006/54/EC.

242 This separates EU law from US law, where the notions of ‘motive’ and ‘intent’ are central to the finding of what is known as disparate treatment.

243 As shown in section 2.4 below, however, this is likely to be a challenge.

244 Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law* C-303/06 EU:C:2008:415.

245 Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law* C-303/06 EU:C:2008:415.

does not follow from [...] Directive 2000/78 that the principle of equal treatment which it is designed to safeguard is limited to people who themselves have a disability [...].²⁴⁶ 'On the contrary, [...] [t]he principle of equal treatment enshrined in the directive in that area applies not to a particular category of person but by reference to the grounds mentioned in Article 1'.²⁴⁷ This finding is important in the context of algorithmic discrimination because it means that direct discrimination can potentially extend to some cases of algorithmic proxy discrimination and miscategorisation. For example, in the present case, direct discrimination could cover cases of algorithmic profiling where users are classified within a protected category although they do not share that characteristic themselves, but because of their proximity with the protected group or a person sharing the protected trait. This type of behavioural discrimination by association could occur where search and click data has been collected about an online user, for instance where such data reveals that the user is interested in whether given restaurants or museums are accessible to wheelchair users. Algorithmic profiling would perhaps result in the user being classified as disabled himself or herself, where in actual fact the user is not but shares his or her life with someone who is. Such classification errors linked to the use of behavioural data as proxy in algorithmic profiling could result in scenarios where the user is denied given opportunities by an algorithm because of his or her misclassification as disabled. Thus, the Court's approach in *Coleman* is a valuable extension of the concept of direct discrimination in the context of algorithmic discrimination.

In light of this approach, the concept of direct discrimination should arguably also extend to situations where individuals or groups are discriminated against because of a characteristic they are assumed or perceived to have, even if this is not the case in reality. This scenario is different from the discrimination by association because in this case it does not involve another individual who indeed shares the protected characteristic, but instead involves discriminatory ascriptions that are not necessarily founded in fact.²⁴⁸ In other words, a person should not need to share a protected characteristic to be recognised as a victim of direct discrimination based on that ground.²⁴⁹ Such situations of discrimination by perception, by ascription or by assumption are particularly relevant in relation to algorithmic discrimination because in many cases algorithmic profiling will result in ascribing given traits or characteristics to people or groups based on, for example, behavioural data, while these inferences might not necessarily be correct (e.g. because of lack of data, lack of granularity or profiling errors). The European Commission has promoted such an interpretation by expressing the opinion that the Racial Equality Directive and the Equal Treatment Directive 'also prohibit a situation where a person is directly discriminated against on the basis of a wrong perception or assumption of protected characteristics'.²⁵⁰ However, this interpretation has not been consistently applied by the Court of Justice in cases of direct discrimination.²⁵¹ For example, in *Kaltoft*, a childminder was perceived as disabled by his employer because of his obesity and, during a redundancy round, he was nominated for dismissal partly on that ground.²⁵² Although the employee himself did not consider himself disabled, the employer's perception that the employee was impaired by his obesity led to his dismissal. Instead of recognising that it was the perception of obesity as a disability on the side of the employer that resulted in differential treatment and *in fine* discrimination, the Court of Justice mandated the national court to find whether such an impairment existed in reality, stating that it would be the condition for obesity to count as a disability and ultimately for a finding of discrimination based on

246 Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law* C-303/06 EU:C:2008:415 [38].

247 Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law* C-303/06 EU:C:2008:415.

248 For a similar distinction, see e.g. Waddington, L and Broderick, A (2018), *Combatting disability discrimination and realising equality: A comparison of the UN Convention on the Rights of Persons with Disabilities and EU equality and non-discrimination law* (European Commission and European Network of Legal Experts in Gender Equality and Non-Discrimination).

249 Such a condition would result in an essentialising approach to non-discrimination law, see e.g. Pothier, D (2001), 'Connecting grounds of discrimination to real people's real experiences' 13 *Canadian Journal of Women and the Law* 37.

250 European Commission (2014), *Joint Report on the application of Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin ('Racial Equality Directive') and of Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation ('Employment Equality Directive')* COM(2014) 2 final (Brussels), 10 available at: <https://eur-lex.europa.eu/procedure/EN/1041602>.

251 The CJEU has so far only interpreted the concept of indirect discrimination as covering situations where given perceptions, even when wrong, result in the discriminatory treatment of individuals. See Judgment of 16 July 2015, '*CHEZ Razpredelenie Bulgaria*' AD v *Komisija za zashtita ot diskriminatsia* C-83/14 EU:C:2015:480.

252 See Judgment of 18 December 2014, *Fag og Arbejde (FOA) v Kommunernes Landsforening (KL)* C-354/13 EU:C:2014:2463[24].

disability.²⁵³ This restrictive application of the concept of direct discrimination in cases where protected grounds are ascribed, perceived or assumed is problematic in light of algorithmic profiling techniques that target given behaviours, interests and affinities and make inferences about people's identities and social memberships on that basis. If direct discrimination does not extend to such situations, it might be difficult to capture situations where algorithmically ascribed identities lead to the differential and disadvantageous treatment of individuals and groups.

Beside this uncertainty regarding the scope of the concept of direct discrimination, even bigger questions arise regarding the relevance of the concept in the context of algorithmic discrimination. As discussed in section 1.2.2, machine-learning algorithms are used to discover patterns in big datasets that combine a variety of variables. Such variables might be wholly unrelated to protected grounds or may only be proxies for protected categories. It has been argued that protected grounds themselves will not usually be used as inputs for such algorithms, hence direct discrimination might be less likely to occur in algorithmic decision making when compared to 'traditional' human decision making.²⁵⁴ Another explanation for the unsuitability of the concept of direct discrimination is that the treatment of data and its categorisation by algorithms might not be cognisable by human brains.²⁵⁵ The variables and categories that an algorithm relies on might not mean anything to humans at all, for instance if they are mere mathematical probabilities. It would therefore be difficult to know whether they can be considered to stand for protected grounds. In addition, the use of variables and the categorisation of data in machine-learning algorithms is in constant evolution as the model learns. Because these categories are not static, it would be difficult to know whether they relate to protected categories at given points in time. One would need to watch the algorithmic model and the way the statistical model used treats the available data over time in order to find out whether unfavourable treatment arises, a task which might not be possible in light of existing accessibility issues.²⁵⁶ As explained in section 1.4.4, due to the 'black box' nature of certain algorithms and the proprietary IP regime underpinning certain business models, this opacity is likely to be a problem in and of itself.²⁵⁷ Categorising algorithmic discrimination as direct discrimination is therefore likely to be a challenge given the opacity of particular algorithms, especially in light of the need to establish a comparator under EU law. Indeed, if the lack of transparency of the functioning of an algorithm prevents the gathering of evidence on how the algorithm has treated or would have treated a group that does not share the protected characteristic at stake (the comparator group), then a finding of direct discrimination might be precluded altogether. Section 2.4 further examines these evidentiary issues.

Furthermore, the likelihood of direct discrimination occurring as unfavourable treatment might diminish in the context of algorithmic discrimination. First, some commentators have suggested that, as awareness about relevant legal obligations increases, direct discrimination is likely to decrease in the developing phase of algorithms.²⁵⁸ Secondly, direct discrimination might diminish in the context of algorithms since the direct input of protected categories for decision making might yield lesser predictive accuracy.²⁵⁹ Thus,

253 Judgment of 18 December 2014, *Fag og Arbejde (FOA) v Kommunernes Landsforening (KL)* C-354/13 EU:C:2014:2463 [62]-[64] and for a critique of this approach, see Waddington, L and Broderick, A (2018), *Combatting disability discrimination and realising equality: A comparison of the UN Convention on the Rights of Persons with Disabilities and EU equality and non-discrimination law* (European Commission and European Network of Legal Experts in Gender Equality and Non-Discrimination) 11, 13, 93.

254 See e.g. Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law' 55 *Common Market Law Review* 1143, 1152.

255 See e.g. Leese, M (2014), 'The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union' 45 *Security Dialogue* 494.

256 See also section 1.4.4.

257 See e.g. Pasquale, F (2015), *The Black Box Society: The Hidden Algorithms Behind Money and Information* (Harvard University Press).

258 See Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law', 1152.

259 Ibid.

developers aware of these risks might remove protected categories from the pool of available variables for algorithmic decision making in order to avoid direct discrimination.²⁶⁰

In light of the above, it can be concluded that the concept of direct discrimination offers conceptual strengths given the lack of relevance of intent and its extension to situations of discrimination by association. However, existing uncertainties regarding its applicability to cases of discrimination by ascription and proxy discrimination as well as doubts regarding the overlap between the law's static categorical approach to disadvantage and algorithms' dynamic and mathematical approach to the categorisation of data might lead to a diminishing relevance and conceptual grasp of this central notion of EU gender equality and non-discrimination law in the context of algorithmic discrimination.

2.3.2 Indirect discrimination: a better conceptual fit with a wide pool of potential justifications

The second central notion of EU gender equality and non-discrimination law is that of indirect discrimination. Although originally absent from EU law, it was developed by the CJEU in its early gender equality case law in the 1980s, taking inspiration from the US equality doctrine.²⁶¹ Now inscribed in the directives, indirect discrimination is defined as situations 'where an apparently neutral provision, criterion or practice would put [members of a protected category] at a particular disadvantage compared with other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary'.²⁶² Instead of focusing on the unfavourable treatment of given groups and individuals because of a given protected ground, the notion of indirect discrimination places the focus on the disadvantageous effects of any given – apparently neutral – practice or measure.²⁶³ Like its direct counterpart, this second concept presents a number of strengths, but also weaknesses, in the context of capturing algorithmic discrimination.

First, as for direct discrimination, the presence or absence of any intent to discriminate is irrelevant. Regardless of whether the developers of an algorithm, the company using an algorithm for commercial purposes, or the administration relying on an algorithm for decision-making purposes intend to discriminate, if they end up disproportionately disadvantaging a protected group, this situation can be captured by the notion of indirect discrimination. Even in the case of what has been termed the 'masking' of direct discrimination²⁶⁴ or, in other words, the concealing of direct discrimination through the use of a proxy for a given protected category in a given algorithmic model, the notion of indirect discrimination will capture what is known as covert direct discrimination.²⁶⁵

Secondly, the concept of indirect discrimination is better suited to the very logic of algorithmic discrimination because, instead of focusing on the treatment of various individuals based on their group membership, it shifts the focus towards the effects of any decision, measure or policy in terms of disadvantage experienced by protected groups.²⁶⁶ As algorithmic discrimination arises from the mining

260 It has been argued that the existence of deep-seated discriminatory biases and the overwhelming under-representation of women and minority groups in STEM and IT-related fields seem to temper this hypothesis; see e.g. Noble, S (2018), *Algorithms of oppression: how search engines reinforce racism* (New York University Press). Additionally, as further addressed in section 2.4, removing protected characteristics from the pool of data available to an algorithm might be problematic as it might lead to concealing discrimination altogether.

261 See e.g. Tobler, C (2005), *Indirect Discrimination: A Case Study into the Development of the Legal Concept under EC Law* (Social Europe Series, Hart Publishing). In terms of case law, see Judgment of 31 March 1981, *JP Jenkins v Kingsgate (Clothing Productions) Ltd* C-96/80 EU:C:1981:80 and Judgment of 13 May 1986, *Bilka – Kaufhaus GmbH v Karin Weber von Hartz* C-170/84 EU:C:1986:204.

262 Article 2(2)(b) Directive 2000/43/EC; Article 2(2)(b) Directive 2000/78/EC; Article 2(b) Directive 2004/113/EC; Article 2(1)(b) Directive 2006/54/EC.

263 See Tobler, C (2008), *Limits and potential of the concept of indirect discrimination* (European Commission).

264 Barocas, S and Selbst, A (2016), 'Big Data's Disparate Impact' 104 *California Law Review* 761.

265 On covert direct discrimination, see e.g. Besson, S (2008), 'Gender discrimination under EU and ECHR law: Never shall the twain meet?' 8 *Human Rights Law Review* 647.

266 See Robin-Olivier, S (2012), 'L'émergence de la notion de discrimination indirecte: évolution ou révolution?' (The emergence of the concept of indirect discrimination: evolution or revolution?) in Fines, F, Gauthier, C and Gautier, M (eds), *La non-discrimination entre les européens* (Non-discrimination among Europeans) (Pedone).

of large datasets, it concerns population groups who share common characteristics. The theoretical underpinnings of the notion of indirect discrimination thus better fit the operation of algorithms than that of direct discrimination. In addition, because indirect discrimination would place the focus on the discriminatory effects of algorithms rather than on their operations, it would offer a way to get around the difficulties exposed in section 1.4 (and further elaborated in section 2.4) in relation to the access to the content of algorithms, the understanding of their operations by non-experts, and questions of liability along the chain of algorithmic decision making, from developers to end users.

Thirdly, the concept of indirect discrimination as interpreted by the CJEU is capable of adequately addressing situations of proxy discrimination where decisions are made on the basis of characteristics related to, but different from, protected grounds. In *CHEZ*, for example, the CJEU implicitly recognised that residency could be a proxy for ethnicity in a case where residents of an area mostly inhabited by Roma people were prevented from accessing their electricity consumption meters by a company that held racist stereotypes against the Roma population.²⁶⁷ The company had decided to place electricity meters out of reach to prevent electricity theft in the area, a policy which it did not pursue in other areas of service provision. The capacity of the concept of indirect discrimination to address proxy discrimination is important in the context of algorithmic discrimination. As demonstrated in section 2.2, even where protected grounds are excluded from the pool of variables of a given algorithm, proxies of such grounds can be used, resulting in similar discriminatory effects. Capturing these situations as indirect discrimination therefore safeguards the effectiveness of EU gender equality and non-discrimination law.

In addition, the concept provides a safety net for tackling proxy discrimination when there is doubt as to whether the link between a given proxy and a given protected ground is direct enough for direct discrimination to arise.²⁶⁸ This is also shown in the CJEU's judgment in *Jyske Finans*, where the Court considered whether a credit institution directly discriminated against a Danish citizen born in Bosnia and Herzegovina by asking him to provide additional proof of his identity compared to Danish citizens born in Denmark. First, relying on the decision in *CHEZ*, the Court recognised that 'the concept of "ethnicity" has its origin in the idea of societal groups marked in particular by common nationality, religious faith, language, cultural and traditional origins and backgrounds'.²⁶⁹ Acknowledging that 'as the list begins with the words "in particular", it is not exhaustive and it cannot therefore be ruled out that a person's country of birth might be included among those criteria', the Court established a link between the applicant's country of birth and his ethnicity. However, it refused to consider the first ground as a proxy for the second, explaining that a person's country of birth 'is only one of the specific factors which may justify the conclusion that a person is a member of an ethnic group and is not decisive in that regard'.²⁷⁰ This led the CJEU to finding no differential treatment based on a proxy for ethnic origin.²⁷¹ Indeed, by holding that 'a person's country of birth cannot, in itself, justify a general presumption that that person is a member of a given ethnic group such as to establish the existence of a direct or inextricable link between those two concepts', the Court cast doubt on the capacity of the notion of direct discrimination to tackle proxy discrimination.²⁷² This is concerning in light of the demonstrated use of residency data and postcodes by algorithms as a way of inferring people's ethnicity.²⁷³ However, the concept of indirect discrimination might apply to capture the group disadvantage at stake, even though, in a controversial move, the Court considered that it was

267 Judgment of 16 July 2015, '*CHEZ Razpredelenie Bulgaria*' AD v *Komisija za zashtita ot diskriminatsia* C-83/14 EU:C:2015:480.

268 Judgment of 6 April 2017, *Jyske Finans A/S v Ligebehandlingsnævnet*, acting on behalf of *Ismar Huskic* C-668/15 EU:C:2017:278. A comparable line of reasoning is developed in *Achbita*, see Judgment of 14 March 2017, *Samira Achbita and Centrum voor gelijkheid van kansen en voor racismebestrijding v G4S Secure Solutions NV* C-157/15 EU:C:2017:203.

269 Judgment of 16 July 2015, '*CHEZ Razpredelenie Bulgaria*' AD v *Komisija za zashtita ot diskriminatsia* C-83/14 EU:C:2015:480 [46] and Judgment of 6 April 2017, *Jyske Finans A/S v Ligebehandlingsnævnet*, acting on behalf of *Ismar Huskic* C-668/15 EU:C:2017:278 [17].

270 Judgment of 6 April 2017, *Jyske Finans A/S v Ligebehandlingsnævnet*, acting on behalf of *Ismar Huskic* C-668/15 EU:C:2017:278 [18].

271 Judgment of 6 April 2017, *Jyske Finans A/S v Ligebehandlingsnævnet*, acting on behalf of *Ismar Huskic* C-668/15 EU:C:2017:278 [25].

272 Judgment of 6 April 2017, *Jyske Finans A/S v Ligebehandlingsnævnet*, acting on behalf of *Ismar Huskic* C-668/15 EU:C:2017:278 [20].

273 See e.g. Williams, B, Brooks, C and Shmargad, Y (2018), 'How algorithms discriminate based on data they lack: challenges, solutions, and policy implications' 8 *Journal of Information Policy* 78.

not applicable in the specific circumstances at hand in *Jyske Finans*.²⁷⁴ This was later confirmed by the CJEU in *Maniero*, a case concerning the award of a scholarship for law students in Germany.²⁷⁵ Award of the scholarship was conditional on holding the German 'First State Examination' in law and the applicant argued that this requirement amounted to indirect discrimination on grounds of ethnic or racial origin because it had the effect of 'placing people of foreign ethnic origin with an equivalent diploma acquired abroad at a disadvantage'.²⁷⁶ While this was not the Court's finding, it confirmed that the concept of indirect discrimination is well-suited to capturing group disadvantages.²⁷⁷

Fourthly, the notion of indirect discrimination is apt to redress proxy discrimination even in situations where the wronged group or individual does not possess the relevant protected characteristic. Indeed, in *CHEZ*, the CJEU replicated the *Coleman* decision prohibiting 'discrimination by association', which concerned direct discrimination, in cases of indirect discrimination.²⁷⁸ Even though the applicant in *CHEZ*, Ms Nikolova, a local shop keeper, was not of Roma origin herself, the CJEU held that the 'principle [of non-discrimination] is intended to benefit also persons who, although not themselves a member of the [protected] group concerned, nevertheless suffer less favourable treatment or a particular disadvantage on one of those grounds', citing *inter alia* the *Coleman* decision.²⁷⁹ This interpretation, which extends the scope of the concept of indirect discrimination to situations of discrimination by association (and by extension discrimination by assumption), offers a further guarantee in relation to redressing proxy discrimination arising from algorithmic profiling.

The above has shown the increasing relevance of the concept of indirect discrimination in the context of algorithmic discrimination,²⁸⁰ both substantively and instrumentally. This is true in particular if the notion is compared to that of direct discrimination. In addition to the issues discussed in section 2.3.1, while it is relatively easy to filter out protected grounds to avoid direct discrimination, it might not be feasible to do so with their proxies so that discrimination might take place despite these precautions. Indeed, societal and structural discrimination may creep into the operation of algorithms if the data used to program or train them reflects biases and stereotypes that have crystallised into patterns of inequality over time.²⁸¹ If debiasing strategies are not put in place and the data is not 'cleaned', it will inevitably reflect structural forms of inequality that originate from the institutionalisation of past discrimination over the course of history. If used by algorithms as training material, the patterns of inequality engrained in this data will be further reproduced, reified and performed by algorithms in their output. The operation of algorithms, because of their reliance on *de facto* biased social data, thus increases the likeliness of occurrences of indirect discrimination. The robust doctrine developed by the CJEU around indirect discrimination could help address this issue and therefore, from a substantive perspective, the notion has significant advantages compared to that of direct discrimination. Instrumentally, moreover, it has been argued that indirect discrimination could be understood as 'a purely instrumental device that assists in the enforcement and expands the scope of the law of direct discrimination', for example by helping 'to

274 Judgment of 6 April 2017, *Jyske Finans A/S v Ligebehandlingsnævnet, acting on behalf of Ismar Huskic* C-668/15 EU:C:2017:278 [33]. For a critique, see Farkas, L (2017), *The meaning of racial or ethnic origin in EU law: between stereotypes and identities*, European Network of Legal Experts in gender equality and non-discrimination (European Commission).

275 Judgment of 15 November 2018, *Heiko Jonny Maniero v Studienstiftung des deutschen Volkes eV* C-457/17 EU:C:2018:912.

276 Judgment of 15 November 2018, *Heiko Jonny Maniero v Studienstiftung des deutschen Volkes eV* C-457/17 EU:C:2018:912 [21].

277 Judgment of 15 November 2018, *Heiko Jonny Maniero v Studienstiftung des deutschen Volkes eV* C-457/17 EU:C:2018:912 [46]-[50].

278 Judgment of 16 July 2015, '*CHEZ Razpredelenie Bulgaria*' AD v *Komisia za zashtita ot diskriminatsia* C-83/14 EU:C:2015:480 and Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law* C-303/06 EU:C:2008:415.

279 Judgment of 16 July 2015, '*CHEZ Razpredelenie Bulgaria*' AD v *Komisia za zashtita ot diskriminatsia* C-83/14 EU:C:2015:480 [49] and [56].

280 In this sense, see also Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law' 55 *Common Market Law Review* 1143; Zuiderveen Borgesius, F and Poort, J (2017), 'Online price discrimination and EU data privacy law' *Journal of Consumer Policy*.

281 See e.g. Mayson, S (2018), 'Bias In, Bias Out', University of Georgia School of Law Legal Studies Research Paper No 2018-35. See also Ridgeway, C (2014) 'Why status matters for inequality' 79 *American Sociological Review* 1.

overcome problems of proving direct discrimination’ or to ‘enable the selective protection of other groups [than those captured by protected grounds] under the guise of discrimination law’.²⁸²

The concept of indirect discrimination and its doctrinal application, however, also entail practical difficulties. While a finding of direct discrimination excludes justification apart from a closed and restricted list of exceptions,²⁸³ establishing a *prima facie* case of indirect discrimination opens a wide pool of possible justifications. The directives provide that no indirect discrimination is to be found where the implicated ‘provision, criterion or practice is objectively justified by a legitimate aim, and the means of achieving that aim are appropriate and necessary’.²⁸⁴ Once a *prima facie* case of indirect discrimination has been established, the burden of proof shifts onto the defendant, which strengthens the position of the applicant.²⁸⁵ This provision then opens up the possibility for a defendant to invoke any justification and put it to the consideration of a court. It translates into a proportionality test, the aim of which is for a court to find out whether the existence of a disproportionate disadvantage can be justified by a measure serving a legitimate interest. Further conditions for that measure to be accepted as justification for an existing disadvantage are its objective, effective and proportionate nature, and its necessity, that is the absence of any other measure that could fulfil the same aim and that would be less detrimental to the wronged group. The grasp of the concept of indirect discrimination is therefore not as stringent as that of the concept of direct discrimination.

The openness of the indirect discrimination test poses a number of problems in relation to algorithmic discrimination. If cases of algorithmic discrimination fall ‘by default’ into the indirect discrimination category, leading to an open pool of possible justifications, legal certainty for potential victims, developers and users of algorithmic systems will decrease as the appreciation of the validity of potential justifications would exclusively be bestowed upon courts. In particular, the application of the ‘necessity’ part of the objective justification test by courts poses questions in light of the trade-off between accuracy and performance on the one hand and non-discrimination on the other that might arise in cases of algorithmic discrimination. Finally, the very boundaries between direct and indirect discrimination might become increasingly blurred in cases of algorithmic discrimination.²⁸⁶ On the one hand, doubts might arise regarding the possibility of directly relating proxy data to any protected ground thus putting in jeopardy the applicability of the notion of differential treatment based on a protected ground. On the other hand, the doctrinal distinction will be undermined if a growing number of cases of algorithmic discrimination fall within the indirect discrimination category ‘by default’ for lack of transparency on the input data used by the algorithmic model.

2.4 Questions of proof, responsibility and liability

As explained in sections 1.3 and 1.4.6, many different players are involved in the planning, development, and use of algorithms. This fragmentation of the algorithmic decision-making chain raises questions about who is responsible for, and who should be held liable, in the event of discrimination arising from the use of algorithms. Should the IT engineers and programmers who build and train these algorithms, i.e. the software company commercialising the algorithmic end product, be held responsible? Is it the entity responsible for selling or providing training datasets to the software company (e.g. because the data it used to train the algorithm was not ‘clean’)? Is it the user of the algorithm, i.e. the entity ordering the

282 Collins, H and Khaitan, T (2018), ‘Indirect discrimination law: controversies and critical questions’ in Collins, H and Khaitan, T (eds), *Foundations of Indirect Discrimination Law* (Hart Publishing) 7.

283 Such exceptions include, for instance, genuine and determining occupational requirements, which can be invoked as laid out in Article 4 Directive 2000/43/EC; Article 4 Directive 2000/78/EC; and Article 14(2) of Directive 2006/54/EC. Article 6 of Directive 2000/78/EC also contains a number of exceptions to direct age discrimination.

284 Article 2(2)(b) Directive 2000/43/EC; Article 2(2)(b) Directive 2000/78/EC; Article 2(b) Directive 2004/113/EC; Article 2(1)(b) Directive 2006/54/EC.

285 Article 8(1) of Directive 2000/43/EC; Article 10(1) of Directive 2000/78/EC; Article 9(1) of Directive 2004/113/EC; Article 19(1) of Directive 2006/54/EC.

286 For an extended analysis, see Xenidis, R, ‘Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law’ (on file with the author).

creation of given algorithms and deploying them for various decision-making purposes (e.g. the human resources department of a private company or an advertising platform)? The questions of responsibility and liability for discriminatory algorithmic decision making are complex and not easily answered.

In addition, this fragmented algorithmic decision-making chain is further complicated when discrimination arises from a technology that integrates various algorithms and combines them with enabling technologies (e.g. the internet of things). As explained in section 1.2.4, AI applications are often complex and made up of various algorithmic and data-generating components. For example, algorithmic decision making can involve situations where the output of one particular algorithm, which itself relies on the data generated by a given connected object (e.g. voice assistants such as Google's Alexa, children's toys or connected home appliances such as washing machines, heaters, etc.), is used as input for another algorithm.

Such a situation creates manifold risks and problems. If one of the connected systems fails, for example because of a technical failure or a misinterpretation of data, this may have the effect that the other systems also fail – resulting in a process of cascading failures.²⁸⁷ Moreover, the interconnectedness of technologies and the fragmented nature of algorithmic decision-making processes multiplies the number of actors involved and makes the distribution of responsibility and liability even more obscure. If discrimination arises at the end of the chain, how can responsibility be traced? Will all those involved face collective liability? Should one particular person or organisation bear the liability burden alone? If so, who should that be? Moreover, when there is a 'human in the loop' during the actual decision-making phase, should the human bear responsibility, should it be the machine, or should it be both? This could produce hybrid liability situations that equality and non-discrimination law might not yet be fit to address.²⁸⁸

Furthermore, these questions arise in the context of another challenge that was uncovered in Chapter 1, namely that of the transparency and explainability of algorithms when one of their defining features is opacity (see section 1.4.4). First, algorithms used for commercial purposes are often proprietary, that is, they are protected under intellectual property law and trade secrets. At the moment, there is no obligation for a company to give access to the algorithms it uses or disclose their characteristics. Secondly, machine-learning algorithms and, in particular, deep-learning algorithms (but even in some cases rule-based algorithms relying on complex decision trees) are often 'black boxes': their inner workings are not intelligible to lay persons, sometimes not even to experts, which makes it difficult to trace and isolate the source of a given discriminatory output. Which correlation is problematic? Does the problem come from biased data or does the algorithmic design reflect a harmful stereotype? Moreover, even if the source of the problem can be pinpointed, machine-learning algorithms are in constant evolution and consequently, the source of discrimination can change over time or disappear altogether.

Specifically in relation to equality and non-discrimination law, the opacity, explainability and transparency challenges might pose a problem in relation to issues of proof, evidence and redress. Not only might it be difficult for data scientists to identify the source of algorithmic discrimination, but any evidence of such discrimination in the algorithmic procedure itself might have disappeared once the algorithm has 'mutated'. In addition, even if data scientists are able to pinpoint the source of a discriminatory algorithmic output, it might be another challenge to make it intelligible to non-experts such as victims, judges, legislators, etc. While the concept of indirect discrimination might provide a way to bypass these difficulties by eliminating the need to open the 'black box' – by focusing on any discriminatory effects rather than the cause of such disadvantage – further challenges related to issues of proof and evidence complicate the task of establishing discrimination. In this light, many discussions regarding responsibility and liability for faulty algorithms focus on different notions ranging from transparency (imposing an obligation to make the content of an algorithm accessible) to accountability (imposing an obligation for those involved in the planning, development and use of algorithms to be accountable) and from

287 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek* (Legal aspects of algorithms that make decisions. An exploratory study) (WODC/Utrecht University Utrecht) section 2.2.

288 On the role of human in the loop and related challenges, see section 1.4.1.

interpretability (imposing an obligation to enable a human to understand the cause or reason of a given algorithmic decision) to explainability (imposing an obligation to enable humans to understand why an algorithm is producing a given outcome and how this outcome has been obtained).²⁸⁹

Another question that arises in this context relates to the burden of proof and the opportunity for victims to identify the existence of discrimination in the first place. Under EU law, the prevailing non-discrimination model is an ex-post adjudicative adversarial system based on individual litigation. This means that the burden of uncovering discrimination, starting proceedings and bringing a case to court lies with the victim of discrimination (or, when allowed under national law, organisations with a legitimate interest).²⁹⁰ This is problematic in light of the isolation of potential victims of algorithmic discrimination.²⁹¹ For example, an online user who has been discriminated against by an internet platform in relation to the price of given products on grounds of her profile (including her gender, cultural affinities, age, etc.) might not even be aware that such discrimination is happening, especially in the context of targeted advertising and personalised services. In this respect, it should be recalled that the CJEU held in *Meister* that there is no right to information for the alleged victim of discrimination, which further complicates the applicant's task of establishing a *prima facie* case of discrimination.²⁹² In *Meister*, the applicant alleged discrimination in relation to her sex, age and ethnic origin in relation to a recruitment process and sought to obtain information on who the other applicants were from the recruiting company. The latter refused to provide such information, thus *de facto* preventing the applicant from establishing a *prima facie* case of discrimination. While the CJEU indicated that the company's refusal could count towards the establishment of such a *prima facie* case of discrimination, it also rejected the existence of a right to information for the applicant. Hence, even in cases that qualify as indirect discrimination where there is no need to open the algorithmic 'black box' but only to provide *prima facie* evidence that discrimination might be happening and where the burden of proof shifts to the defendant, it might be difficult for individual victims, and even monitoring bodies as the case may be, to gather the necessary evidence in the first place. Similarly, even when *prima facie* evidence can be gathered, it might be a challenge for judges to operate the necessary proportionality and objective justification test in cases of indirect discrimination. As pointed out in the previous section, it might be difficult to assess whether a software company responsible for the discriminatory output of an algorithm could have opted for a less impactful solution in light of the highly technical explanations the company is likely to present in relation to the trade-offs that it must make.

2.5 Conclusion

This chapter has shown how the inconsistencies, ambiguities, and shortcomings of EU gender equality and non-discrimination law inevitably also limit its capacity to deal with algorithmic discrimination. Many of these issues existed before the development of algorithms and have been identified by scholars for a long time. However, the increasing use of algorithms in all domains of society sharpens these problems and multiplies the risks that discrimination falls in the cracks of the EU legal framework. This might be the case in relation to discrimination on grounds of age, disability, sexual orientation and religion in the area of goods and services, where profiling algorithms are increasingly used by advertisers and commercial providers. This might also be the case in relation to existing exceptions to EU gender equality law in the fields of the media, advertising and education. Overall, the limitations in the EU legislation's material scope clearly diminish the effectiveness of EU non-discrimination legislation in helping to eradicate

289 See Miller, T (2017), 'Explanation in artificial intelligence: Insights from the social sciences', arXiv Preprint arXiv:1706.07269 and Molnar, C (2020), *Interpretable machine learning: a guide for making black box models explainable* available at: <https://christophm.github.io/interpretable-ml-book/>, Chapter 2.

290 In this sense, see Judgment of 10 July 2008, *Centrum voor gelijkheid van kansen en voor racismebestrijding v Firma Feryn NV* C-54/07 EU:C:2008:397; Judgment of 25 April 2013, *Asociația Accept v Consiliul Național pentru Combaterea Discriminării* C-81/12 EU:C:2013:275; and Judgment of 23 April 2020, *NH v Associazione Avvocatura per i diritti LGBTI — Rete Lenford* C-507/18 EU:C:2020:289.

291 Xenidis, R and Senden, L (2020), 'EU non-discrimination law in the era of artificial intelligence: mapping the challenges of algorithmic discrimination', in Bernitz, U and others (eds), *General Principles of EU Law and the EU Digital Order* (Alphen aan den Rijn, Kluwer Law International).

292 Judgment of 19 April 2012, *Galina Meister v Speech Design Carrier Systems GmbH* C-415/10 EU:C:2012:217.

algorithmic discrimination. To the extent that national non-discrimination laws have a broader material scope or have not implemented available exceptions, they offer better possibilities for preventing and redressing algorithmic discrimination.

Beyond the material scope, algorithmic discrimination also raises issues in relation to the exhaustive list of protected grounds that form the substance of the personal scope of EU equality law. Due to the proxies and correlation challenge discussed in section 1.4.3, it is questionable whether this provides the best possible legal avenue for protection against algorithmic discrimination. Algorithms change the nature of discrimination, increasing risks of miscategorisation based on users' behavioural data, proxy discrimination as well as intersectional discrimination due to algorithms' granular profiling abilities. It will be rare for an algorithm to discriminate only or decisively on the basis of a protected ground, since it will usually base its output on a multitude of different factors and variables that are all statistically correlated. First, this raises the question of whether a ground like 'gender' can be defined sufficiently broadly as to cover a range of proxy grounds. Secondly, it means that, in many cases, the output of an algorithm will be based on a combination of characteristics and behaviour that is unique to a particular (group of) person(s). The focus on a few protected grounds and the lack of legal recognition of notions of multiple and, in particular, intersectional discrimination in the current EU and national legislation means that such instances of 'combined' or highly differentiated discrimination cannot be effectively redressed.

Moreover, the issue of algorithmic discrimination to some extent reconfigures the traditional divide between direct and indirect discrimination. The problem of algorithmic discrimination does not neatly fit the traditional doctrinal paradigms of EU and national non-discrimination law. It has been pointed out in the literature on algorithmic discrimination that direct discrimination is unlikely to arise because it is improbable that designers of algorithmic systems directly input protected grounds as negative variables in supervised algorithmic models.²⁹³ Legal scholars have also claimed that biases in the processing of data would lead to indirect rather than direct forms of discrimination.²⁹⁴ A scholarly consensus has thus emerged that the concept of indirect discrimination would be a better conceptual fit for algorithmic discrimination than that of direct discrimination.²⁹⁵ As demonstrated in this chapter, these arguments deserve to be nuanced because algorithmic discrimination blurs the boundaries between the doctrines of direct and indirect discrimination, leading to a classification challenge.²⁹⁶ Given that it might be increasingly difficult to identify differential treatment based on protected grounds in the context of algorithmic operations, the notion of indirect discrimination might become a conceptual 'refuge' to capture the discriminatory wrongs of algorithms. However, such a trend might increase legal uncertainty given the open-endedness of the objective justification test applicable in this case.

Finally, the enforcement of equality law poses its own problems. Risks that algorithmic discrimination goes unnoticed and unredressed are high in light of the lack of transparency and explainability of algorithmic decision making, the highly fragmented and hybrid chain of responsibility in algorithmic decision making and the related challenge of determining where liability lies, as well as the difficulties in providing proof and evidence. In addition, the individual rights-based model of EU equality law is problematic because many of the issues related to algorithmic bias and discrimination are of structural nature (e.g. socio-historical biases in data). Due to the problems of transparency and evidence discussed above, it will also be difficult for individual victims to establish that they are disadvantaged or harmed by the use

293 Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law' 55 *Common Market Law Review* 1143, 1151.

294 Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law', 1152.

295 See e.g. Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law', 1143; Zuiderveen Borgesius, F (2020), 'Strengthening legal protection against discrimination by algorithms and artificial intelligence', *International Journal of Human Rights* 1; Barocas, S and Selbst, AD (2016) 'Big data's disparate impact', *California Law Review* 104, 671; Zarsky, T (2017) 'An analytic challenge: discrimination theory in the age of predictive analytics' 14 *Journal of Law and Policy for the Information Society*, 11.

296 See Xenidis, R (2020), 'Tuning EU equality law to algorithmic discrimination: Three pathways to resilience' 27(6) *Maastricht Journal of European and Comparative Law* 736 and Xenidis, R, 'Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law' (on file with the author).

of algorithms. Consequently, equality and non-discrimination systems that are strongly based on an individual rights and individual complaints system, and lack opportunities for bringing collective or class actions, may not be able to adequately address the problems and challenges inherent to algorithmic decision making. To correct the various gaps in the legal framework highlighted in this chapter, several proposals can be made, which will be further elaborated in Chapter 4.

3 Challenges for the European states in relation to algorithmic discrimination

3.1 Examples of the use of algorithms in European countries

3.1.1 Introduction

Algorithms are actively used in many public and private sectors all over Europe. In the public sector, uses range from labour market policy to predictive policing and from education to transportation and immigration. In all these fields, algorithms are used mainly for three purposes: the allocation of resources, the assessment of risks and the detection of irregularities (e.g. fraud, illegal usages, etc). In the private sector, algorithms are often used to enhance efficiency, speed and precision of decision making as well as for purposes such as price-setting, targeting, and bringing supply and demand together.

In this chapter, based on input provided by national experts, first some relevant examples of the (projected) use of algorithms in the public sector are presented, in particular in labour market policy, social welfare, education, policing and fraud detection, the administration of justice, and media regulation (section 3.1.2). In section 3.1.3, a number of examples from the private sector are discussed, specifically in relation to employment decisions, banking and insurance, targeted advertising, pricing and retail, renting and tourism. The examples provided in this section are not intended to be exhaustive or representative (algorithms are, for instance, also often used in healthcare and diagnostics and the market for short-term holiday rentals of houses and apartments), but they may help to give an idea of the diversity of possible uses and the variety of aims and objectives that are pursued by the use of algorithms. The examples further serve the function of offering a more tangible basis for explaining some of the major equal treatment and non-discrimination challenges that are seen to arise as a result of such uses, which are discussed in the other sections of this chapter.

3.1.2 Examples of the use of algorithms in the public sector

3.1.2.1 Labour market policy

In several countries, predictive profiling algorithms are used (or are projected to be used) by Government agencies and other public bodies to support their labour market policies.²⁹⁷ This is often done with the objective of being able to identify and predict the job opportunities for certain unemployed persons or estimate their need for training. In **Austria**, for instance, the Labour Market Service (AMS) has developed an algorithm that, based on previous statistical labour market data, can be used to determine future labour market chances of job applicants.²⁹⁸ The algorithmic prognosis can help determine the assignment of job applicants to one of three pre-defined groups. More expensive resources such as one-on-one job counselling and access to job training programmes are then allocated in considerably higher measure to persons in the first two groups, who would be supposed to have better chances in the labour market. The third group contains persons whose labour market chances are defined as low by the algorithm, and who will be offered a different type of support in accordance with their profile.²⁹⁹

297 In some countries, the use of algorithms in relation to the labour market has yet to be seen, but this may change in the (near) future; see e.g. **Germany**: Fröhlich, W and Spiecker genannt Döhmann, I (2018), 'Können Algorithmen diskriminieren?' (Can algorithms discriminate?) (*Verfassungsblog*, 26 December 2018) available at: verfassungsblog.de/koennen-algorithmen-diskriminieren/.

298 The launch was originally planned for mid-2020, but has been postponed to 2021.

299 Ongoing research project 'An Inquiry into the AMS algorithm from a socio-technical perspective' by the Technical University of Vienna (TU Wien) and the Federal Chamber of Labour, running until October 2020; see further <https://www.oew.ac.at/ita/projekte/der-ams-algorithmus>.

Already in use by the public employment service of Flanders (VDAB), in **Belgium**, is a similar AI program, which can help predict the chance that an unemployed job seeker will *not* find a job within the next six months. The objective of this is to improve the identification of job seekers in need of personalised support.³⁰⁰

Another example of this type of algorithm can be seen in **Poland**, where the Ministry of Labour and Social Policy has introduced a system based on profiling the unemployed to decide on how to distribute labour market programmes.³⁰¹ On the basis of the result, the system can assign the respondents to one of three profiles differing with regard to the degree of assessed readiness to start work and the type of assistance from the employment office.

The use of algorithms in employment policy can further be illustrated by the online application 'My employability' used by the **Croatian** Employment Service (CES). This application has been developed to help job seekers calculate the probability of finding employment within the next 12 months, based on their replies to a set of questions that correspond to a number of given parameters.³⁰² The parameters include county of residence, age, sex, age of the youngest child,³⁰³ applicability of special measures (e.g. for Croatian war veterans or persons with disabilities), work experience, level of education, field of education, unemployment history, reason for ending the previous job (e.g. dismissal or expiry of a fixed-term contract) and previous occupation. The probability is calculated according to input and is expressed in percentage points as a graphical display of two columns of probability: one for the specific designated county and one depicting the average probability at the level of Croatia. The results as shown by the app to the user highlight that work experience, level and field of education, history of unemployment and the reasons for ending the previous job are the most important indicators for the calculation. The app also invites users to check what happens when some of the input data is changed, e.g. occupation, level of employment, previous work experience, etc. Contrary to other algorithmic applications described above, which are used to decide on resource allocation, the Croatian app is not intended to affect the person's position on the labour market in any way.

3.1.2.2 Social welfare

Predictive algorithms are sometimes applied in relation to social welfare issues, but this use is (still) limited and not many examples have been reported. For **Finland**, an analysis has been made of data on all child welfare clients in 2002-2016 and the whole local population's health and social data in a big Finnish city in order to find factors that would predict marginalisation of certain persons or groups.³⁰⁴ The research found 280 factors that predict that a child will be in need of child welfare services. These factors and results can now be used in the allocation of preventive resources. In **Spain**, a programme for smart social home care has been developed for predicting the social aid needs of the elderly.³⁰⁵ According to the developers, the system aggregates data about social services, health, population, economic activity, utility usage, waste management, and more and uses this data to identify and predict groups and areas that will need urgent help.

300 Information based on interview by national expert Nathalie Wuiame with G Vanhumbeeck (Director for Innovation), K Scheerlinck (AI team leader) and V Buekenhout (Data Protection Manager), VDAB, online Teams meeting (12 June 2020).

301 Škop, M and others (2019), *alGOVrithms. State of Play. Report on Algorithms Usage in Government-Citizens Relations in Czechia, Georgia, Hungary, Poland, Serbia and Slovakia* (Krzysztof Izdebski ed, Fundacja ePaństwo) available at: epf.org.pl/pl/wp-content/uploads/sites/2/2019/05/alGOVrithms-State-of-Play-Report.pdf, 38.

302 The system is available at: <https://stapweb.hzz.hr/>.

303 Possible options are 0-2, pre-school, school or 15+.

304 Koskenvuo, K (2019), 'Tekoäly avuksi nuorten syrjäytymisen ehkäisemisessä' (AI helpful in preventing youth marginalisation), *Sosiaalivakuutus*, 14 May 2019, available at: <https://sosiaalivakuutus.fi/tekoaly-avuksi-nuorten-syrjäytymisen-ehkäisemisessä/>.

305 See <https://bismart.com/en/business-intelligence-solutions/smart-social-home-care-for-aging-population/>. According to the website, the system is being used in Bilbao and Barcelona.

3.1.2.3 Education

In some countries, algorithms are used in relation to education. This is exemplified by the Parcoursup system that was introduced in **France** for university admissions, as already mentioned in section 2.1.3.³⁰⁶ The system was used to allocate places in higher education institutions to incoming students. It raised concerns in relation to the use of certain income and residency data and was also criticised for its lack of transparency.³⁰⁷ The system randomly drew candidates for admission to universities when selective credentials were not required to avoid ranking. **Poland** also makes extensive use of this function of algorithms, for instance for the assignment of children to nurseries, kindergartens and high schools, as well as for admission to colleges. For example, in assigning children to nurseries, the city of Wrocław has introduced an algorithm that relies on data obtained from the parents' declarations: the number of children in the family, a disability certificate, whether parents were employed or in education, the place of residence in Wrocław, and the age of the child.³⁰⁸ Based on this information it can be calculated whether children are qualified to access specific nurseries.³⁰⁹

3.1.2.4 Policing and fraud detection

Notable use of (predictive) algorithms is made in policing and the detection of fraud, for instance in relation to taxation or social benefits. Sometimes these algorithmic applications are still at a pilot stage.³¹⁰ In the **Netherlands**, for example, several pilots are being run where data derived from using sensing technology is combined with predictive algorithmic analyses to detect risk behaviour, e.g. for pickpocketing or mobile 'banditism'.³¹¹ In **Lithuania**, police representatives have expressed their positive attitude towards the idea of preventing criminal acts by means of (i) automated analysis of data gathered by surveillance cameras located in different cities and (ii) the usage of object recognition algorithms, as well as the use of biometric data in voice, fingerprint, palm print and iris recognition systems.³¹²

Already in use in the **Netherlands** is the criminality anticipation system (CAS).³¹³ This system can be used by the police to predict the risk of crimes being committed on the basis of an algorithmic analysis of data on crime reports. The analysis can help to identify 'hot spots' and 'hot times' that can be used to step up police presence and indicate interventions. In **Germany**, police authorities in the state of Hesse and North Rhine-Westphalia use a program called 'Gotham', which is provided by Palantir, a US-based private company.³¹⁴ The Gotham program is used for 'predictive policing', in particular for preventive detection of

306 For more information, see Administrative Guideline 24 April 2017 cancelled by the Conseil d'Etat (*Circulaire 24 avril 2017 annulée par le Conseil d'Etat*), 22 December 2017, Association SOS Education, Promotion et défense des étudiants et Droits des Lycéens Nos. 410561, 410641, 411913.

307 Although the code supporting the national algorithm was made public, local algorithms for the allocation at university level were not and the system was criticised for this lack of transparency. See e.g., Défenseur des Droits (2019), 'Parcoursup: le Défenseur des droits, dans deux décisions, recommande plus de transparence, de mobilité et de mixité et des mesures d'accompagnement adaptées pour les personnes handicapées' available at: <https://www.defenseurdesdroits.fr/fr/communiqu%C3%A9-de-presse/2019/01/parcoursup-le-defenseur-des-droits-dans-deux-decisions-recommande-plus>.

308 For more information, see <https://wroclife.pl/wroclife-poleca/rekrutacja-do-zlobkow/>; <https://wroclaw.wyborcza.pl/wroclaw/7,35771,23544016,zamieszczenie-z-rekrutacja-do-zlobkow-299-dzieci-najpierw-sie.html>.

309 Initially, the system appeared to make mistakes in classifying children on the borders of different age groups, which resulted in incorrect assignments, but the error was detected and the algorithm has been improved. <https://wroclife.pl/wroclife-poleca/rekrutacja-do-zlobkow/>.

310 This also appears to be the case in **Norway**; for more information, see the report by the Norwegian Board of Technology: Teknologirådet (2015), *Forutseende politi. Kan dataanalyser hjelpe politiet til å være på rett sted til rett tid?* (Foresighted police. Can data analysis help the police to be in the right place at the right time?) available at: https://teknologiradet.no/wp-content/uploads/sites/105/2018/05/ForebyggendeAnalyse_endelig_WEB.pdf.

311 The **Netherlands**: e.g. Eindhoven University of Technology (2018) 'Zakkenrollers herkennen dankzij data-correlatie' (Recognising pickpockets thanks to data correlation), (11 July 2018), available at: www.tue.nl/nieuws/nieuwsoverzicht/zakkenrollers-herkennen-dankzij-data-correlatie.

312 <http://utena.policija.lrv.lt/lt/naujienos/dirbtinio-intelektu-taikymas-ateities-policijos-veikloje>.

313 Netherlands Police (2017), 'Criminaliteits Anticipatie Systeem verder uitgerold bij Nationale Politie' (Criminality anticipation system further used by National Police) (16 May 2017) available at: www.politie.nl/nieuws/2017/mei/15/05-cas.html; see also Jansen, F (2018), 'Data driven policing in the context of Europe', ERC-funded project 'Data Justice: Understanding datafication in relation to social justice' (DATAJUSTICE) starting grant (2018-2023), Working Paper 7 May 2018, 8 available at: www.datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-Data-Driven-Policing-EU.pdf.

314 www.fr.de/politik/diese-datenanalysen-koennen-diskriminierung-verstaerken-13498141.html.

possibly dangerous persons or situations. The program can, for example, determine whether a suspicious person has connections with so-called ‘endangerers’, based on information on, for example, the fact that these persons have stayed in the same house, have had cell phone contact, or even have sat in the same car during a police check.³¹⁵ **Spain** is active in this field with applications such as an algorithm used by the police to evaluate the risk of women reporting gender violence (VioGén),³¹⁶ a tool to spot false reports made to the police (VeriPol) and an application that helps to predict recidivism (e-Riscarvi).³¹⁷ In a similar vein, the Durham Police in the **United Kingdom** uses a harm assessment risk tool to predict the risk of reoffending, using information such as postcodes (and possibly ethnicity data),³¹⁸ and the South Wales Police made use of a predictive profiling algorithm based on automatic facial recognition.³¹⁹

In **France**, some algorithms are in use to help detect social security fraud,³²⁰ as well as fight tax evasion, as part of the ‘Openfisca’ system.³²¹ In the **Netherlands**, the system risk indication (SyRI) allows a predictive algorithm to search the data of residents in certain municipalities for patterns that could indicate social security fraud, although the system is currently being revised to make it more privacy-proof.³²² The state educational loan fund in **Norway** uses machine learning in order to discover fraud and cheating on student’s living allowances.³²³ In **Poland**, the clearance chamber ICT system (STIR) enables the exchange of information between banks and the National Tax Administration with the objective of combating VAT fraud.³²⁴ Financial data are derived from banks and cooperative savings and credit unions to conduct analyses of operations in order to determine whether account holders perform certain types of actions that indicate that they may be using their bank accounts for illegal activity. The STIR algorithm determines the risk indicator, which constitutes the central premise according to which the head

315 For criticism of the programme, see e.g. Herberg, R and Lindhoff A (2020) ‘Automatisch verdächtig: Polizei setzt zunehmend auf umstrittene US-Software’ (Automatically suspicious: Police increasingly rely on controversial US software), *Frankfurter Rundschau*, 24 January 2020, available at: www.fr.de/politik/hessen-umstrittene-polizei-software-palantir-automatisch-verdaechtig-13454012.html.

316 <https://algorithmwatch.org/en/story/viogen-algorithm-gender-violence/>.

317 See the inventory made by the Observatory of Algorithms with Social Impact (OASI), established by the Eticas Foundation, <https://eticasfoundation.org/algorithms/>.

318 Rovatsos, M, Mittelstadt, B and Koene, A (2019), *Landscape Summary: Bias in Algorithmic Decision-Making. What is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it?* (Centre for Data Ethics and Innovation) available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf.

319 On the SWP algorithm, see *R (Bridges) v CCSWP and SSHD* [2019] EWHC 2341 (Admin).

320 Guidelines using algorithms to predict the probability of fraud for social benefits given to families, recommend cherry-picking files of people who were born outside of Europe (CNAF No. 2012-142 of 31 August 2012, 41 recommande notamment de ‘cibl[er] les personnes nées hors de l’Union européenne’; see Défenseur des Droits (2017), ‘Lutte contre la fraude aux prestations sociales : à quel prix pour les droits des usagers?’ (Fighting social benefits fraud: what price for users’ rights?), 20, available at: https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=16746; see also Lettre-circ. CNAF No. 2012-42 ‘Modifications apportées par la loi de financement de la sécurité sociale pour 2012 en matière de fraude – Plan national de lutte contre la fraude’ (Amendments made by the Social Security Financing Act for 2012 with regard to fraud – National plan to combat fraud) (August 2012) 41. See further, National Delegation to Combat Fraud (Délégation Nationale à la Lutte contre la Fraude) (2014), ‘Le “data mining”, une démarche pour améliorer le ciblage des contrôles’ (Data mining: an approach to improve targeting investigations) (Paris, 14 January 2014), available at: www.economie.gouv.fr/files/files/DNLF/fichier_data_mining_joint.pdf, 3.

321 Ministerial order by Ministry of Finance of 28 August 2017 creating an automated anti-fraud system entitled ‘fraud targeting and request enhancement’ (*Arrêté du 28 août 2017 modifiant l’arrêté du 21 février 2014 portant création par la direction générale des finances publiques d’un traitement automatisé de lutte contre la fraude dénommé « ciblage de fraude et valorisation des requêtes*).

322 Decision of 1 September pertaining to the amendment of the SUWI Decree in relation to rules for combatting fraud by means of data exchange and the effective use of data known to the Government using SyRI (*Besluit van 1 september 2014 tot wijziging van het Besluit SUWI in verband met regels voor fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekend zijnde gegevens met inzet van SyRI*), *Staatsblad* (Official Journal) 2014, 320, available at: <https://zoek.officielebekendmakingen.nl/stb-2014-320.html>; see also Letter of 23 April 2020 to the President of the House of Representatives by the State Secretary for Social Affairs and Employment, Tamara van Ark, on a court judgment regarding SyRI (*Kamerbrief van 23 april 2020 naar aanleiding van vonnis rechter inzake SyRI, Kamerstukken II* (Parliamentary Papers) 2019/20, 17050, No 593 available at www.rijksoverheid.nl/documenten/kamerstukken/2020/04/23/kamerbrief-naar-aanleiding-van-vonnis-rechter-inzake-syri).

323 Jørgenrud, M (2018), ‘Lånekassen tar i bruk kunstig intelligens for å avdekke juks’ (The State Educational Loan Fund is using artificial intelligence to detect cheating in students exams), *Digi.no*, (22 January 2018), available at: www.digi.no/artikler/lanekassen-tar-i-bruk-kunstig-intelligens-for-a-avdekke-juks/426299.

324 Škop, M and others (2019), *alGOVrithms. State of Play. Report on Algorithms Usage in Government-Citizens Relations in Czechia, Georgia, Hungary, Poland, Serbia and Slovakia* (Krzysztof Izdebski ed, Fundacja ePaństwo), available at: epf.org.pl/pl/wp-content/uploads/sites/2/2019/05/alGOVrithms-State-of-Play-Report.pdf, 36.

of the National Tax Administration may request a block on the bank account of a given entity.³²⁵ Finally, local authorities in the **United Kingdom** are allowed to voluntarily adopt risk-based verification (RBV) in relation to housing benefits and council tax benefits.³²⁶ The RBV works by assigning a risk rating to each applicant for such benefits, which then determines the level of identity verification required.³²⁷ This allows the local authority to target and focus resources on ‘... those cases deemed to be at highest risk of involving fraud and/or error’.³²⁸ Someone with a high-risk rating might be subject to additional checks, visits and an increased requirement to provide documentation.³²⁹

3.1.2.5 Administration of justice

Algorithms are not often used in courts as yet, although there are some experiments with algorithm-supported judicial decision making in the **Netherlands**.³³⁰ Sometimes, moreover, algorithms are used to predict how judges will decide their cases. In **France**, for example, algorithmic applications have been developed to anticipate the future decisions of judges in civil cases and their allocation of remedies according to past behaviour. Using software like ‘Predictice’ or ‘Supralegem’, employers or other decision-makers can try to avoid liability by knowing how judges will decide their future cases.³³¹ In other countries, algorithms are used in the judiciary in a more administrative and organisational manner. An example of this can be found in **Poland**, where the algorithm-based system of random allocation of cases (SLPS) assigns cases to judges of the particular court on a once-per-day basis. This system has been implemented in all 364 ordinary courts.³³²

3.1.2.6 Media regulation

It is well known that many media platforms make use of algorithms to identify and remove hate speech and other forms of discriminatory, insulting or defamatory expressions. **France** has recently adopted legislation to introduce a system that would allow the tracking of users of the internet who are responsible for hate speech, although this legislation has now been declared partially unconstitutional.³³³ In **Spain**, a tool to detect hate speech on Twitter has been developed with the help of the National Bureau for the Fight against Hate of the Ministry of Home Affairs.³³⁴

325 See Škop, M and others (2019), *alGOVrithms. State of Play*, 36; <https://epf.org.pl/pl/2019/05/13/maszyny-sluzbie-panstwa-algorytmy-wplywaja-zycie-obywateli-obywatelek-polsce-wybranych-krajach-regionu>.

326 See Allen, R (2020), ‘Artificial intelligence, machine learning, algorithms and discrimination law: the new frontier’ (Discrimination Law Conference, Michael Rubenstein Conferences, London, 31 January 2020) available at: https://482pe539799u3ynseg2hl1r3-wpengine.netdna-ssl.com/wp-content/uploads/2020/02/Discrimination-Law-in-2020.FINAL_-1.pdf.

327 Allen, R (2020), ‘Artificial intelligence, machine learning, algorithms and discrimination law: the new frontier’, para 74; see also Housing Benefit and Council Tax Benefit Circular, HB/CTB S11/2011.

328 Allen, R (2020), ‘Artificial intelligence, machine learning, algorithms and discrimination law: the new frontier’.

329 Ibid.

330 Van der Put, M (2019), ‘Kan artificiële intelligentie de rechtspraak betoveren?’ (Can Artificial Intelligence enchant the judicial branch?) 2 *Rechtstreeks* 50.

331 Bourcier, D and De Filippi, P (2018), ‘Transparence des algorithmes face à l’open data: quel statut pour les données d’apprentissage?’ (Transparency of algorithms in light of open data: what is the status of learning data?) 167 *Revue française d’administration publique* (3) 525.

332 Tarkowski, A (2019), ‘Report. Automating Society. Poland’ *AlgorithmWatch*, (29 January 2019) available at: <https://algorithmwatch.org/en/automating-society-poland/>; Škop, M and others (2019), *alGOVrithms. State of Play. Report on Algorithms Usage in Government-Citizens Relations in Czechia, Georgia, Hungary, Poland, Serbia and Slovakia* (Krzysztof Izdebski ed, Fundacja ePaństwo) available at: epf.org.pl/pl/wp-content/uploads/sites/2/2019/05/alGOVrithms-State-of-Play-Report.pdf, 14. No details about the algorithm are known and the source codes are kept secret, which makes it difficult to know how the algorithm works and what data are used as input.

333 Constitutional Council: ‘Bill to fight against hate speech on internet, platforms and search engines in 24 hours adopted by the National Assembly 13 May 2020 but cancelled in most part by the Constitutional Council as an infringement of freedom of expression considering the difficulty to monitor all content closely’, Conseil Constitutionnel, Decision No. 2020-801 DC of 18 June 2020, available at: www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm; see further Defender of Rights and CNIL (National Data Protection Authority of France) (2020), *Algorithmes: prévenir l’automatisation des discriminations* (Algorithms: preventing automated discrimination), joint report, 7 available at: www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-en-num-16.07.20.pdf.

334 See the inventory made by the Observatory of Algorithms with Social Impact (OASI), established by the Eticas Foundation, <https://eticasfoundation.org/algorithms/>.

3.1.3 Examples of use of algorithms in the private sector

Just as in the public sector, algorithms are often relied on in the private sector. As this subsection illustrates, the purposes of algorithmic applications are more varied than in the public sector, where they are mostly used as aids to decision making and as tools for risk assessment, detection and efficient allocation of resources. Particularly noticeable is the use, in two cases, of so-called emotional AI, that is, systems that are able to detect human emotions through analysing facial expressions and traits.

3.1.3.1 Employment and platform work

Algorithms can be very convenient in making employment decisions, and indeed, there appears to be a growing trend in their use by private companies for this purpose. In **Finland**, for example, it was estimated in 2017 that some form of AI was used in 40 000 recruitments per year, and this number is assumed to be increasing quickly.³³⁵ Similarly, in **Poland**, the Panoptykon Foundation has presented an analysis according to which more and more recruitment processes are becoming automated.³³⁶ This means that a machine is responsible for pre-selecting and rejecting some of the candidates who do not meet the pre-defined criteria, e.g. in terms of language skills, type of education or experience counted by years of work. The company Amazon Fulfillment Poland has been reported to use various algorithmic tools in this respect.³³⁷ In terminating employment contracts, for example, Amazon relied on an algorithmic indication of the percentage share of sickness absence in the company's working time.³³⁸

Another employment-related use of algorithms in **Poland** concerns the practice adopted by the (state-controlled) bank PKO BP of 'collecting smiles'.³³⁹ This is a system incorporating individual sensors, combined with an advanced algorithm, to count a consultant's smiles during conversations with clients. The system's creators assume that the more an employee smiles at work, the more satisfied the customer will be, which in return would motivate the employee.

In relation to platform work, there is currently a case pending in **Italy** on multinational food delivery companies using the 'Frank' algorithm. According to the trade unions that have brought the case, 'the algorithm, in elaborating the reputation rankings of the cyclists, which in fact determine future job opportunities and booking priorities for deliveries, marginalises, to the point of excluding them [...] those who do not manage to be available to log into the work slots assigned to them; riders who do not adapt to the logic of the algorithm are gradually excluded from work opportunities, leading in some cases to their being logged out by the system'.³⁴⁰ In other words, riders whose availability does not allow them to accept all the rides proposed by the algorithm are disadvantaged in subsequent allocation of work and are at risk of being completely excluded from work opportunities.

335 Nurmilaakso, T (2017), 'Tekoäly valitsee jo Suomessakin parhaat työntekijät – Tutkijat: "Lisää syrjintää, seksismiä ja rasismia"' (Artificial intelligence in Finland already selects the best employees – Researchers: "Increasing discrimination, sexism and racism"), *Yle Finnish Broadcasting Company*, (7 October 2017) available at: <https://yle.fi/aihe/artikkeli/2017/10/07/tekoaly-valitsee-jo-suomessakin-parhaat-tyonhakijat-tutkijat-lisaa-syrjintaa>.

336 Styczyński, J (2019), 'Algorytmy zamiast działu kadr? RODO studzi zapalę' (Algorithms instead of HR? RODO dampens down the enthusiasm), *Gazeta Prawna* (11 July 2019) available at: <https://serwis.gazetaprawna.pl/praca-i-kariera/artykuly/1421586,rodo-a-automatyczne-systemy-rekrutacji-pracownikow.html>.

337 See <https://www.pb.pl/jak-to-jest-z-dyskryminacja-algorytmiczna-964121>.

338 The media reports that the company has now abandoned the practice. See further e.g. Matłacz, A (2018), 'Nie można zwalniać za częste L4. Amazon przegrał w sądzie' (One can't lay off for frequent use of L4. Amazon lost in court) *Wyd. WKP, Sekcja: Kadry i BHP, WKP*, (10 December 2018), www.prawo.pl/kadry/zwolnienie-z-pracy-za-czeste-przebywanie-na-l4-wyrok-so-ws,342357.html.

339 <https://choice.pl/7,156282,25239406,smile-for-PLN-5-million-algorithm-evaluates-if-employee-bp-sufficiently.html>.

340 Rassegna Sindacale (2019) 'Si chiama "Frank", l'algoritmo che discrimina' (Its name is Frank, the algorithm that discriminates) (18 December 2019), available at: www.rassegna.it/articoli/si-chiama-frank-lalgoritmo-che-discrimina.

3.1.3.2 Banking and insurance

Predictive algorithms can be helpful in making creditworthiness assessments and defining actuarial risk groups, which makes them an attractive tool in the banking and insurance sectors. For example, in **Lithuania**, there are experiments with a peer-to-peer insurance platform, Ooniq.³⁴¹ According to its privacy policy, ‘based on information provided to the system’, this platform may engage in automated, algorithm-driven calculation of the amount of payment for lost or damaged devices such as mobile phones or tablets. The data that is used to give input to the system includes the model of the device and the amount of damages suffered.

In the banking sector, it has been reported that many banks in **Belgium**,³⁴² **Estonia**,³⁴³ **France**,³⁴⁴ **Germany**,³⁴⁵ **Poland**, and **Spain**³⁴⁶ use profile setting for access to bank credits and similar services.

3.1.3.3 Targeted advertising, price-setting and retail

The use of algorithms is increasingly popular in the sectors of retail and advertising. In particular, predictive algorithms and profiling can be used for distributing personalised or targeted advertisements, especially on social media or websites, and they can also be relied on to engage in individualised or group-based price setting. In **Denmark**, for example, RockCrew, an organisation supplying temporary workers to stage crews, showed their ad on Facebook to men only.³⁴⁷ In **Germany**, the Berlin Public Transport Company offered targeted discounts to women on International Women’s Day using facial recognition.³⁴⁸ RIMI, one of the two largest food retailers in **Latvia**, allows a client to swipe his/her client’s card in an information machine to get personalised promotion offers.³⁴⁹ In **Lithuania**, an algorithm is currently in use in a fully automated store called Pixevia, where artificially intelligent systems can tell precisely which customer has taken what product in real time.³⁵⁰ Subsequently, this can result in targeted marketing. Finally, in **Norway**, pizza chain Peppe’s Pizza was shown to use facial recognition and algorithmic analyses to personalise the photos of available offers and options on the restaurant’s digital advertising screen.³⁵¹ Data used to personalise the offers made on the screen were, amongst others, gender, age, appearance and mood (e.g. smiling). The result was that men were often shown pictures of pizza with steak/meat, while women were shown healthy salads.

341 See www.ooniq.lt/.

342 Chancerelle, L (2019), ‘La lutte contre les discriminations en Europe à l’ère de l’intelligence artificielle et du big data’ 381 *Journal du Droit des Jeunes* 25.

343 A list of parameters and a query processing algorithm has been assigned to each service. Queries from the merchant to the bank are directed to the URL www.lhv.ee/banklink; see www.lhv.ee/images/docs/Bank_Link_Technical_Specification-EN.pdf.

344 Gadhoum Y, Gueyie J-P and Siala MK (2007), ‘La décision de crédit: Procédure et comparaison de la performance de quatre modèles de prévision d’insolvabilité’ (The credit decision: procedure and performance comparison of four insolvency forecasting models) 2 (224-225) *La Revue des Sciences de Gestion* 177.

345 The State Minister for Homeland, Local Affairs, Construction and Equality of North Rhine-Westphalia suggested that women are often disadvantaged per se by systems of granting loans; they would have to pay higher interest rates, offer more collateral or would not get a loan in the first place – see www.presseportal.de/pm/30621/4604677.

346 Spielkamp, M (2019), *Automating Society. Taking stock of automated decision-making in the EU* (A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations, 1st edn, January 2019) available at: https://algorithmwatch.org/wp-content/uploads/2019/01/Automating_Society_Report_2019.pdf.

347 The ad is mentioned in a news story available at: <https://www.dr.dk/nyheder/indland/ikke-kvinder-virksomheder-maalretter-jobopslag-paa-facebook-til-maend>.

348 State Anti-Discrimination Body of Berlin (2019), *Algorithmen und das Recht auf digitale Gleichbehandlung. Dokumentation der Fachwerkstatt der Landesstelle für Gleichbehandlung – gegen Diskriminierung* (Algorithms and the right to digital equality. Documentation of the specialist workshop) available at: www.berlin.de/sen/lads/ueber-uns/materialien/. See also www.berlin.de/sen/lads/assets/ueber-uns/materialien/20191216_algorithmen-barrierearmes-pdf.pdf.

349 See supermarket chain RIMI’s website: www.rimi.lv/mans-rimi/mans-rimi-kartes-prieksrocibas.

350 See <https://www.pixevia.com/automated-stores>.

351 Tennøe, T and Valevatn, J (2018), ‘Slik overvåker maskiner følelsene dine’ (How machines control your feelings), *Aftenposten*, (26 February 2018, last updated 28 February) available at: www.aftenposten.no/viten/i/ka8jka/slik-overvaaker-maskiner-foelelsene-dine. This example is also referred to on the Norwegian Data Protection Authority’s website: www.personvernbloggen.no/2020/01/31/nar-reklamen-ser-deg/.

3.2 Problems related to algorithmic decision-making

Section 3.1 has shown that algorithms are used throughout Europe, both in the public and the private sector, and with a variety of different objectives and purposes. As discussed in section 1.4, there are a variety of challenges involved in such (projected) uses of algorithms, in particular if seen from the perspective of equality and non-discrimination. The scale on which algorithms allow for decision making is unprecedented, just like the speed of such decision making. There are also challenges relating to the inherent risk of cognitive bias in the use of algorithms, their opacity and ‘black box’ nature, the strong reliance of algorithms on correlations and proxies, and the complex determination of responsibility for the development and use of algorithms. In addition, Chapter 2 reviewed a number of specific risks, gaps and problems related to equality and non-discrimination, and their regulation, on the EU level.

Before turning to a closer analysis of the national responses to such risks and challenges, it is useful to explain on a more general level the main problems, risks and issues that are associated with algorithmic decision making in the various European countries, with a particular focus on issues related to equality and non-discrimination. Referring to the examples of use discussed in section 3.1, the current section explains a number of algorithm-related risks and problems that have been identified by national experts: biases in data (section 3.2.1), (indirect) discriminatory effects of using algorithms (section 3.2.2), transparency issues (3.2.3), problems related to the detection of discrimination (section 3.2.4), and issues of responsibility (section 3.2.5). Finally, section 3.2.6 provides information on the level of diversity and the existence of a ‘gender digital gap’ in STEM fields in Europe.

3.2.1 Biases in data

In relation to many of the examples of the use of algorithms discussed in section 3.1, national experts have expressed or noted concerns about biases in the data that are used to develop and train the algorithms. For example, it was discussed in section 3.1.2.1 that in **Austria**, an algorithmic application has been developed to predict the labour market chances of unemployed persons and allocate resources accordingly. This algorithmic allocation procedure has been criticised for various reasons, such as its lack of transparency as regards the use of data and its unclear approach towards the reproduction of potential biases, notably in relation to the database used for training and providing input to the algorithm.³⁵² Specifically, critics have pointed out that the use of criteria such as sex, age, caring responsibilities, health issues and citizenship in the algorithm’s decision-making procedure might reinforce existing inequalities on the basis of gender, ethnic origin, disability and age. The reason for this is that labour market statistics show that persons nearing retirement ages, immigrants, persons with educational or language deficits, low-skilled workers, and women with informal care duties already suffer from considerable negative stereotypes in the Austrian labour market. If the statistical data that form the basis of the algorithm are not corrected for this, it is feared that the algorithm itself may be set up to give biased results. The Austrian public employment service has recognised that the algorithm ‘has shown that caring responsibilities affect women’s labour market opportunities but not men’s’, but did not disclose how it would correct that bias.³⁵³ Currently, in response to this and similar criticisms and because of the lack of a clear legal basis for the collection and processing of personal data, the implementation of the programme has been postponed.³⁵⁴

352 National Equality Ombud (Gleichbehandlungsanwaltschaft), inquiry to the Austrian Labour Market Service, 11 March 2019, see (GER), available at: https://www.gleichbehandlungsanwaltschaft.gv.at/dam/jcr:4dd65631-38ae-43e5-8289-6fbd0cc29ef0/Schreiben_an_AMS_Bundesgesch%C3%A4ftsstelle_final_11.03.2019.pdf; see also the response by the AMS: Arbeitsmarktservice Österreich, Nr.: GEZ: BGS/VOR/003/2019 of 25 April 2019, response to inquiry by the National Equality Ombuds. See also Czák, A (2019), ‘Das Problem mit dem AMS-Algorithmus’ (The problem with the AMS-algorithm), *Epicenter.works*, (17 October 2019) available at: <https://epicenter.works/content/das-problem-mit-dem-ams-algorithmus>.

353 Arbeitsmarktservice Österreich, Nr.: GEZ: BGS/VOR/003/2019 of 25 April 2019, response to inquiry by the National Equality Ombuds, see (GER), available at: https://www.gleichbehandlungsanwaltschaft.gv.at/dam/jcr:4dd65631-38ae-43e5-8289-6fbd0cc29ef0/Schreiben_an_AMS_Bundesgesch%C3%A4ftsstelle_final_11.03.2019.pdf.

354 Szigetvari, A (2020), ‘Datenschutzbehörde kippt umstrittenen AMS-Algorithmus’ (Data protection authority overturns controversial AMS algorithm), *Der Standard* (20 August 2020) available at: <https://www.derstandard.at/story/2000119486931/datenschutzbehoerde-kippt-umstrittenen-ams-algorithmus>.

In **Belgium**, there are similar concerns related to the VDAB's AI programme, which predicts the chance that an unemployed job seeker will not find a job within the next six months, with the aim of better identifying job seekers in need of personalised support. In practice, women appeared to be more likely to indicate that they want a temporary job, which has prompted the question of whether the algorithm should take that as a fact (and therefore propose such jobs to women) or should not take this into account.³⁵⁵

In a similar vein, the 'My employability' tool in **Croatia** uses the age of users' children as a parameter that is relevant to employability. Especially when the input is '0-2' years of age, this has been reported to have a negative impact on the predictions on the employability of women, whereas for men the parameter was not even displayed or taken into account. Thus, the algorithm is seen to replicate existing bias and stereotypes. Moreover, even though it is only a tool to calculate probabilities, in the national expert's opinion, it could have the potential effect of discouraging women with small children from (re)entering the labour market.

As a last example, there is the use of risk based verification (RBV) by local authorities in the **United Kingdom** in relation to housing benefits and council taxes. In relation to this system, a first analysis of a small sample has suggested that there could be gender bias in the algorithm, because all those identified by the RBV as 'high risk' in the sample taken were working women, although it also has been remarked that further studies would be needed to confirm the suggestion.³⁵⁶

3.2.2 Discriminatory effects

Mostly as a result of in-built biases and stereotypes, national experts report that algorithms can easily cause direct and indirect discrimination. More specifically, algorithms can sometimes lead to sexist decisions, as was observed by the High Council on Equality of **France** in its report on sexism in the use of algorithms by the media and on the internet.³⁵⁷ In addition, it was mentioned that in **France** personalised price-setting for goods and services could cause discrimination by raising prices for gendered products for menstruation.³⁵⁸ Likewise, in **Germany**, the example has been given of the Berlin Public Transport Company offering targeted discounts to women on International Women's Day, using facial recognition.³⁵⁹ This particular use relied on a binary distinction between men and women and was strongly based on stereotypes. It has been reported that personalised behavioural advertising in the real estate market in **Italy** can have the effect of discriminating according to ethnicity and social class, thus contributing to creating 'new ghettos' on the one hand, and luxury districts on the other.³⁶⁰ Another problem in **Italy** is that profiling through personal data can lead to the determination of credit ratings or insurance premiums according to lifestyle or driving, or the inclusion in certain 'social clusters' based on ethnic or geographical origin.³⁶¹ Finally, the fraud detection system, SyRI, in the **Netherlands** might have a discriminatory

355 Information based on interview by national expert Nathalie Wuiame with G Vanhumbeeck (Director for Innovation), K Scheerlinck (AI team leader) and V Buekenhout (Data Protection Manager), VDAB, online Teams meeting (12 June 2020).

356 See Allen, R (2020), 'Artificial intelligence, machine learning, algorithms and discrimination law: the new frontier' (Discrimination Law Conference, Michael Rubenstein Conferences, London, 31 January 2020) available at: https://482pe539799u3ynseg2hl1r3-wpengine.netdna-ssl.com/wp-content/uploads/2020/02/Discrimination-Law-in-2020-FINAL_-1.pdf. It is also mentioned, however, that this could be a coincidence, or the algorithm might have 'learnt' a discriminatory correlation.

357 Haut Conseil à l'Égalité entre les femmes et les hommes (2019), *1er état des lieux du sexisme en France* (1st survey of sexism in France), report No. 2018-01-07 STER 038, available at: www.haut-conseil-egalite.gouv.fr/IMG/pdf/hce_etatdeslieux-sexisme-vf-2.pdf.

358 Allodocteurs (2019) 'Pourquoi les règles coûtent cher?' (Why are periods expensive?) (14 January 2019, last updated 15 January 2019) available at: www.allodocteurs.fr/maladies/gynecologie/pourquoi-les-regles-coutent-cher_26235.html.

359 State Anti-Discrimination Body of Berlin (2019), *Algorithmen und das Recht auf digitale Gleichbehandlung. Dokumentation der Fachwerkstatt der Landesstelle für Gleichbehandlung – gegen Diskriminierung* (Algorithms and the right to digital equality. Documentation of the specialist workshop) available at: www.berlin.de/sen/lads/ueber-uns/materialien/. See also www.berlin.de/sen/lads/assets/ueber-uns/materialien/20191216_algorithmen-barrierearmes-pdf.pdf.

360 Olivi, G (2019), 'Algoritmi che discriminano, ecco perché serve l'approccio giuridico' (Algorithms that discriminate, here is why we need a legal approach), *Network Digital 360*, (4 June 2019) available at: www.agendadigitale.eu/cultura-digitale/algoritmi-che-discriminano-ecco-perche-serve-lapproccio-giuridico/.

361 Olivi, G (2019), 'Algoritmi che discriminano, ecco perché serve l'approccio giuridico'.

and stigmatising effect to the extent that the system led to heightened supervision in certain ‘problem neighbourhoods’.³⁶²

3.2.3 Transparency problems and lack of information

Another problem that is often mentioned on the national level is closely related to the transparency and explainability challenges discussed in section 1.4.4: algorithms are notoriously difficult to comprehend and, in many cases, little or no transparency is provided as to the data used as input or for training. Indeed, in the **Netherlands**, the unavailability of detailed information on the SyRI algorithm for fraud detection was one of the main reasons that the district court of The Hague ruled that it was incompatible with fundamental rights law.³⁶³ The district court criticised the system’s lack of transparency and the difficulties involved in auditing its operation and effects. Moreover, the court considered that the risk models used by the algorithm were kept secret for operational reasons. As a consequence, it could not be assessed whether they had discriminatory effects, which the court held to be unacceptable.

Secrecy is also an issue in relation to the system of random allocation of cases in the judiciary in **Poland**. There, not only the source code of the algorithm used by this system is subject to secrecy, but also the principles of its operation, and even the results generated.³⁶⁴ This makes it very difficult to understand how the algorithm works and whether it may entail elements of bias and discrimination.

To such transparency and secrecy issues, the problems of complexity and the black-box nature of algorithms can be added. For example, the national expert for **Norway** remarked that ‘very few people have the qualifications to create complex algorithms ... When it is done commercially, there is no real control with how it is done, while the consumer actually has little idea what they’re actually asking for.’

Finally, there may be other transparency issues involved in the use of algorithms. In **France**, for example, the *Parcoursup* algorithm, which was used to assign students to universities, was heavily criticised because its objectives were not properly defined. For that reason, eventually, the system was annulled by the French Council of State.³⁶⁵ Here, the transparency problems did not so much relate to the algorithm as such or the data used, but to the lack of clarity regarding the reasons and objectives for its use.

3.2.4 Detecting algorithmic discrimination

Even though many algorithms may lead to confirmation of biases or may have discriminatory or stereotyping effects, another major problem that is perceived on the national level relates to the difficulties involved in detecting and identifying algorithmic discrimination. Many national experts mention that it may not always be obvious if an algorithm really is discriminatory or generates discriminatory effects. In 2020, for example, in **Germany** the Conference of the Federal and State Ministers for Equality (GFMK) pointed out that, due to the complexity of the matter, it seemed unrealistic that those affected would be able to detect and pursue algorithmic discrimination.³⁶⁶

Some concrete examples of the problems that can arise in relation to detection of algorithmic discrimination can be seen in **Poland**, where algorithms are used in relation to the assignment of pupils to nurseries, kindergartens and schools.³⁶⁷ Taking into account the type of data collected and the opacity

362 Rechtbank (District Court) The Hague 5 February 2020, ECLI:NL:RBDHA:2020:865, para 6.92.

363 Rechtbank (District Court) The Hague 5 February 2020, ECLI:NL:RBDHA:2020:865, paras 6.90 and 6.95.

364 Izdebski, K (2020), ‘(Nie)słodka tajemnica losowania sędziów’ (The (un)sweet secret of the judges’ draw) *Fundacja ePaństwo*, (13 January 2020) available at: <https://epf.org.pl/pl/2020/01/13/nieslodka-tajemnica-losowania-sedziow/>.

365 Administrative Guideline 24 April 2017 cancelled by the Conseil d’Etat, (*Circulaire 24 avril 2017 annulé par le Conseil d’Etat*) 22 December 2017 Association SOS Education, Promotion et défense des étudiants et Droits des Lycéens Nos. 410561, 410641, 411913.

366 See www.gleichstellungsministerkonferenz.de/documents/20-07-01-schlussprotokoll-der-sonder-gfmk-25_juni-2020_2_3_1595231802.pdf, 18ff.

367 See further section 3.1.2.3.

of the algorithm behind them, it has been observed that it is difficult to identify any discriminatory elements. Possible bias could result from a municipality's assumptions as to which factors are to be taken into account when enrolling children in educational institutions (and what weights should be assigned to them), but it could also be the result of errors in the construction of the algorithm. Moreover, it has been remarked that the eventual decision on the assignment of an individual pupil will be based on the result of a combination of many factors of different weights. This could make it challenging to detect an error or a specific instance of discrimination.

Another illustration is the 'smiles' ranking that is used in **Poland**, which involves a facial recognition system that detects the number of times a consultant smiles during a meeting with a client.³⁶⁸ As the expert for Poland has noted, in the absence of permanent recordings of the signals received by the facial recognition system, post factum verification of the adequacy of the collected data is in practice impossible, which makes discrimination very hard to detect.

3.2.5 Responsibility issues

It has been remarked by national experts that even if a case of (indirect) algorithmic discrimination can be detected, it is not always easy to contest it because of the compound responsibility that is often evident, which is caused by the involvement of many players in the process of planning, developing and using an algorithm. This problem is recognised by many of the national experts and is further discussed in sections 3.3.1.3 and 3.3.2.3 below. Importantly, moreover, it has been pointed out that this challenge may be an even greater one than that of the involvement of many players alone. For **Ireland**, for example, the expert has mentioned that multinationals that use allegedly discriminatory algorithms in recruitment or promotion may have their Europe Middle East and Africa (EMEA) headquarters or European headquarters in Ireland, but the recruitment may take place somewhere else, i.e. in the parent country. It may be difficult to deal with such discrimination as it is not taking place in Ireland under Irish and EU law.

3.2.6 A gender digital gap in European countries

Finally, specific diversity problems have been reported in nearly all countries studied in this report both in STEM-related educational curricula and in STEM-related professional communities. In particular, the IT sector struggles to attract and keep a female workforce. This reflects the wider context: in 2017, only 12 % of AI researchers in the world were female.³⁶⁹ Several interacting explanations for this can be identified, including gender stereotypes, gender segregation in working life, the exclusion of women from IT jobs through discrimination and the lack of role models for female IT workers. The statistics and figures below offer a picture of the situation in the EU and EEA countries to date.

In **Austria**, STEM curricula had a 34 % share of female students in 2016 against, on average, 61 % in other curricula.³⁷⁰ In particular, only 16 % of IT students were female. After their education, the integration in the labour market of IT graduates also differs, with a 94 % figure for males and a lower figure of 88 % for females, as well as a 7 % difference in male and female salaries. Statistics show almost no change in female participation in STEM education since 2007/2008. In **Croatia**, the share of women employed in the ICT sector is 13 %, which is below the EU average of 17.2 %.³⁷¹ In **Denmark**, women

368 See section 3.1.3.1.

369 Nguyen Doan, H T, Briault, C and Moisy, C (2019), 'Artificial intelligence: the future of man?', *Coverstory, Mekkur* (March-April 2019), available at: https://www.cc.lu/uploads/tx_userccpublications/COVER_STORY_OK.pdf, 58.

370 See, Binder, D and others (2017), *MINT an öffentlichen Universitäten, Fachhochschulen sowie am Arbeitsmarkt Eine Bestandsaufnahme* (STEM at public universities, technical colleges and on the labour market, an inventory), Institute for Advanced Studies, Vienna, available at: <https://irihs.ihs.ac.at/id/eprint/4284/1/2017-ihs-report-binder-mint-universitaeten-fachhochschulen.pdf>.

371 European Commission (2019) *Women in Digital Scoreboard 2019*, available at: <https://ec.europa.eu/digital-single-market/en/news/women-digital-scoreboard-2019-country-reports>.

make up only 24 % of IT employees overall.³⁷² In the **Netherlands**, a study found the number of female students in STEM subjects in 2019 slowly increasing, while the number of women in technical positions increased by 24 % between 2013 and 2019.³⁷³ In **Estonia**, in 2019, women represented only 22 % of the ICT sector workforce.³⁷⁴ The expert for Estonia has reported that there is a low share of women among programmers due to widespread acceptance of the myth that programming is a man's field.³⁷⁵ Similar trends are noticed in **Finland**, where concerns have been expressed about the lack of attractiveness of IT education and professions for women and the fact that educational choices are highly gendered.³⁷⁶ **Germany** also faces a significant underrepresentation of women in IT education as less than 20 % of all IT students are female.³⁷⁷ Researchers also report a digital gender gap in **Greece** and note that digital exclusion is reinforced when gender combines with other exclusionary factors such as disability, age, race and socioeconomic background.³⁷⁸ This problem was noted by the Greek Minister of Education in 2020 in an article entitled 'Women in science: A bet that we have to win', in which she attributed the underrepresentation of women in STEM worldwide to gender stereotypes and segregation in the job market, which affect women's career choices and chances, as well as the pay gap, unequal access to funding and unequal family responsibilities.³⁷⁹ Similar concerns have been expressed in **Luxembourg**, which has been reported as having one of the lowest rates of women's participation in IT and STEM sectors.³⁸⁰ The number of female **Polish** IT students is increasing but remains low (14.6 % in 2018–2019). A breakdown of this statistic indicates that women represent 40 % of the student body in IT fields pertaining to data analysis and processing, 32.1 % of the total number of students in computer science and econometrics, and only 8.9 % in industrial computer science studies.³⁸¹ The proportion of women studying IT majors has increased in **Slovakia** from 3–5 % to 10–12 % in the last five years, but IT professions remain insufficiently attractive to women.³⁸² In **Spain**, data shows an important gender gap in ICT related professions and in women's access to ICT studies.³⁸³ For example, only 15.6 % of ICT professionals were women in 2017, and this number is decreasing, while the number of male ICT professionals is growing.³⁸⁴

372 See, Møllergaard, T (2019), 'New IT Women's Council must get more women in IT and tech', HK News (14 June 2019), available at: www.hk.dk/aktuelt/nyheder/2019/06/14/hk-privat-lancerer-nyt-it-kvinderaad.

373 National Technology Pact (2019) 'Techniepact monitor 2019' (Technology Pact Monitor 2019), available at: www.techniepactmonitor.nl/.

374 Kantar Emor (2020), 'Klaasseinad ja klaaslagi Eesti IKT-s: müksud naiste osakaalu suurendamiseks IKT valdkonnas hariduses ja tööturul' (Glass walls and glass ceilings in Estonian ICT: opportunities to increase the share of women in ICT in education and the labour market) available at: www.sm.ee/sites/default/files/klaaslagi_ja_klaasseinad_eesti_ikt-s_vahearuanne_2020_kantaremor.pdf.

375 Ibid.

376 Statistics Finland provides statistics on educational gender segregation. Statistics show that only 20 % of persons with a degree in AI and ITC are women. There is no change in sight, as only 20 % of students in the field were women in 2019, see Statistics Finland (2020), *Koulutusalojen sukupuolenmukainen segregatio* (Gender segregation by field of education), available at: <https://www.stat.fi/tup/tasaarvo/koulutus#segregatio>. Some studies show that employers are actively trying to get more women into the field, but with little success, see e.g. Talouselämä (2018) *Tivin selvitys paljastaa: naisia yhä vähemmistö IT-alalla, mutta moni firma pyrkii saamaan naisia alalle* (Tivi's study shows: women are still a minority in the IT branch, but many firms try to find more women into the field), (11 January 2018), available at: www.stat.fi/tup/tasaarvo/koulutus#segregatio.

377 See, Rähm, J (2019), 'Lack of women in the IT industry', *Deutschlandfunk* (19 January 2019) www.deutschlandfunk.de/allein-unter-maennern-frauenmangel-in-der-it-branche.724.de.html?dram:article_id=438786.

378 Kamberidou, I, and Pascall, N (2020), 'The digital skills crisis: engendering technology—empowering women in cyberspace', *4 European Journal of Social Sciences Studies* (6) 1–33, available at: <http://www.tinyurl.com/ulxl9od>.

379 Kerameos, N (2020), 'Γυναίκες στις θετικές επιστήμες: Ένα στοίχημα που πρέπει να κερδίσουμε' (Women in science: A bet that we have to win) available at: <https://www.tovima.gr/2020/03/10/vimagazino/kerameos-gynaikes-stis-thetikes-epistimes-ena-stoixima-pou-prepei-na-kerdisoume/>.

380 See Interview with Marina Andrieu in *Merkur* (Edition No. 2/2019) 84, available at: https://issuu.com/maisonmoderne/docs/issuu_02_2019_merkur.

381 Fundacja Edukacyjna Perspektywy (2019) *Women at polytechnics (Kobiety na politechnikach)*. The report is based on the publication of the Central Statistical Office of Poland, *Higher education institutions and their finances* (Szkoly wyższe i ich finanse) by POLON, Poland's largest data repository on Polish science and higher education and unpublished data collected by the Educational Foundation 'Perspectives', during many years of cooperation with technical universities throughout Poland; see also www.dziewczynynapolitechniki.pl/pdf/raport-kobiety-na-politechnikach2019.pdf.

382 See Action plan for the digital transformation of Slovakia for 2019–2022 (Government Resolution No. 337/2019, 3 July 2019) 26, available in English at: www.vicempremier.gov.sk/wp-content/uploads/2019/10/AP-DT-English-Version-FINAL.pdf.

383 State Secretary for Digital Advancement (2020), *Libro Blanco de las mujeres en el ámbito tecnológico* (White paper on women in technology), available at: www.mineco.gob.es/stfls/mineco/ministerio/ficheros/libreria/LibroBlancoFINAL.pdf.

384 This seems to be a European trend and not a Spanish particularity; compare Eurostat (2018), 'Digital economy and society statistics', *Statistics Explained*.

In other words, women in ICT account for only 2 % of total female employment in Spain.³⁸⁵ At the source of this problem lies the persistent absence of women in ICT-related education and training. Numbers have been stagnating, with 33 % of women in ICT-related training in 1999 compared to 37.4 % in 2017.³⁸⁶ In technological university and non-university degrees in **Spain**, the gender gap has also been increasing, reaching 12.6 % in 2017. Finally, in the **United Kingdom**, women made up only 16 % of entrants to undergraduate courses in engineering, technology and computer science in 2018-2019, compared to 56 % of total female undergraduate entrants.³⁸⁷ The House of Commons Science and Technology Committee and the House of Lords Committee on Artificial Intelligence have recognised the problem and acknowledged that diversity among professionals developing algorithms is a key tool in tackling algorithmic discrimination.³⁸⁸

The underrepresentation of women and minorities in STEM educational and professional communities has been widely recognised by scholars as problematic. Practical examples provided in the literature include driving machines or weapons that are trained only by men's voices and thus perform less well in recognising female voices, and assistant chatbots called after female names and welcoming/serving robots with feminine appearances, which can perpetuate harmful stereotypical perceptions of women's roles in society and in particular the typical association between female and caring/assisting activities.³⁸⁹ It has also been pointed out by scholars that a lack of diversity in data and design leads to systems biased against underrepresented groups.³⁹⁰

3.3 Awareness of risks of algorithmic discrimination in European countries

3.3.1 Public discussions on the impact of algorithms on gender equality and non-discrimination

3.3.1.1 General discussions on the discriminatory potential of AI

The countries covered by this report can be divided into three groups as regards the degree of public awareness and discussion concerning the specific question of the impact of algorithms on gender equality and non-discrimination. In about a third of the countries, some discussion on this specific topic can be noted as part of, often broader, public debates on AI and ethics or AI and fundamental rights. These discussions are either being led by public authorities and policymakers or by NGOs. In another (slightly larger) set of countries, specific discussions on the impact of algorithms on equality rights have been much more limited, and debates have focused on broader ethical or related legal questions such as transparency, data protection, privacy and the use of algorithms by public authorities. Finally, in the final group of countries, national experts indicate that the issue of algorithmic discrimination has not yet permeated the public sphere to any significant extent.

Thus, a majority of countries have witnessed at least some degree of public discussion on issues of AI and discriminatory risks. In a number of countries, the European Commission's 'White paper on Artificial Intelligence' has triggered discussions in the form of a national consultation process and has given rise to

385 Spanish Association for Digitalization (2018), 'Mujeres en la Economía Digital en España' (Women in the digital economy in Spain).

386 State Secretary for Digital Advancement (2020), *Libro Blanco de las mujeres en el ámbito tecnológico*, 9.

387 <https://www.officeforstudents.org.uk/data-and-analysis/equality-and-diversity/>.

388 House of Commons Science and Technology Committee (2018), *Algorithms in Decision Making (Fourth Report of Sessions 2017-2019)* (HC 351, 15 May 2018) available at: <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf> and House of Lords Select Committee on Artificial Intelligence (2018) *AI in the UK: Ready, Willing and Able?* (HL Paper 100, 16 April 2018) available at: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>.

389 These studies show how stereotypical gender roles are perpetuated in the design of chatbots and robots: machines in charge of caring and assistant roles are designed to be identifiable as 'female'. Devillers, L (2017), *Des robots et des hommes* (Plon). See also www.lemonde.fr/festival/article/2018/09/25/les-etudes-de-genre-se-penchant-sur-le-sexe-des-robots_5359786_4415198.html and see, for more scholarly work on this, Criado Perez, C (2019) *Invisible women: Exposing data bias in a world designed for men* (Random House) and D'Ignazio, C and Klein, LF (2020), *Data feminism* (MIT Press).

390 Criado Perez, C (2019) *Invisible women: Exposing data bias in a world designed for men* (Random House); D'Ignazio, C and Klein, LF (2020), *Data feminism* (MIT Press).

public responses from national policy makers and/or civil society organisations involved in promoting AI ethics, good AI governance and/or protecting human rights in the digital age. For example, in **Denmark**, the Government has initiated a dialogue with a number of stakeholders, including social partners, in the context of the white paper consultation process. Stakeholders have called for clear responsibility in AI systems, non-discriminatory outcomes that respect fundamental rights, transparency in AI design, diversity in the development of AI in order to avoid bias and EU-regulation facilitating access to the reasons behind algorithmic decisions.³⁹¹ The national expert for **France** has reported that the debate has gradually shifted from a focus on data protection, privacy, ethics and infringements of public liberties to greater focus on issues of inequalities and discrimination.³⁹² In **Greece**, the publication of the white paper received considerable attention in the media, where the issue of algorithmic discrimination has been regularly reported on since about 2017.³⁹³ The **Lithuanian** AI Think Tank has been reported as having begun an assessment of existing legal regulations, not restricted to but including issues of algorithmic discrimination, in light of the specific recommendations made by the white paper.³⁹⁴ In **Malta**, the National Commission for the Promotion of Equality organised a 2020 International Women's Day conference on AI and gender equality, which brought together different stakeholders, such as academics and social partners, and covered topics including gender biases and gender mainstreaming in AI.³⁹⁵ In the **Netherlands**, the Rathenau Research Institute reacted to the white paper by advising the Commission 'to uphold fundamental rights and public values' in its AI policy and to '[g]uard against the pitfalls of a risk-based approach' that could risk jeopardising fundamental rights and societal ethics.³⁹⁶ The Dutch Government started the discussion on AI and fundamental rights early on by commissioning several scientific reports on the subject as well as presenting various policy briefs, which then served as a basis

391 See the memorandum to the European Affairs Committee of the Danish Parliament (2020), available at: www.ft.dk/samling/20191/almindel/ERU/bilag/210/2171335/index.htm.

392 Defender of Rights and CNIL (joint report) (2020), *Algorithms: preventing the automation of discrimination (Algorithmes: prévenir l'automatisation des discriminations)*; Villani, C (2018), *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*, (Giving meaning to artificial intelligence: for a national and European strategy), report to the Prime Minister, (8 September 2018) 163, available at: www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf (which sets out the risk of gender and racial discrimination); Institut Montaigne (2020), *Algorithmes: contrôle des biais* s.v.p. (Algorithms: Bias control, please), (March 2020).

393 On reactions to the white paper, see Startupper (2019) 'Η ΕΕ θέτει "υπό έλεγχο" την Τεχνητή Νοημοσύνη' (The EC sets Artificial Intelligence under control), (19 February 2020), available at: <https://startupper.gr/news/57762/i-ee-theti-y-po-elegcho-tin-techniti-noimosyni/>; Euronews (2020) 'Η «λευκή βίβλος» της ΕΕ για την τεχνητή νοημοσύνη' (The European Commission's white paper on artificial intelligence) (19 February 2020) available at: <https://gr.euronews.com/2020/02/19/i-lefki-vivlos-tisee-gia-tin-techniti-noimosini/>; Kathimerini (2020) 'Τεχνητή Νοημοσύνη με κανόνες' (Artificial Intelligence with rules), (25 February 2020) available at: www.kathimerini.gr/1066039/article/epikairothta/kosmos/tehnith-nohmosynh-me-kanones; Kathimerini (2020) 'Η Ευρωπαϊκή μάχη της Τεχνητής Νοημοσύνης' (The European battle of Artificial Intelligence), (20 February 2020) available at: www.kathimerini.gr/1065686/article/epikairothta/kosmos/h-eyrwpaiikh-maxh-ths-tehniths-nohmosynhs; Fortune Greece (2020) 'Τι προβλέπει η Λευκή Βίβλος της Ε.Ε. για την τεχνητή νοημοσύνη' (What does the European Commission's white paper on artificial intelligence provide?) (19 January 2020) available at: www.fortunegreece.com/article/ti-provlepi-i-lefki-vivlos-tis-e-e-gia-tin-techniti-noimosini/. On algorithms and discrimination, some examples include: Economy Today (2019) 'Οι αλγόριθμοι οδηγούν την ανθρωπότητα σε μεγαλύτερες διακρίσεις – Οι ακαδημαϊκοί προειδοποιούν για σοβαρές επιπτώσεις στις κοινωνίες του μέλλοντος' (Algorithms drive humanity to bigger discriminations – Academics warn for serious repercussions on future societies), (12 December 2019) available at: https://economytoday.sigmalive.com/financial-times/13992_oi-algorithmoi-odigoun-tin-anthropotita-se-megalyteres-diakriseis; IN (2017), 'Οι αλγόριθμοι και γιατί χρειάζονται περισσότερο κοινωνικό έλεγχο' (Algorithms and why more social control is needed) (12 February 2017) available at: www.in.gr/2017/02/12/tech/oi-algorithmoi-kai-giati-xreiazontai-perissotero-koinwniko-elegxo/; Ampa-Lifo (2018) 'Η τυφλή εμπιστοσύνη στους αλγόριθμους συντηρεί τις υπάρχουσες διακρίσεις – Οι αλγόριθμοι δεν είναι αντικειμενικοί, είναι γνώμες ανθρώπων ενσωματωμένες σε κώδικα' (Blind trust in algorithms preserves existing discrimination – Algorithms are not objective, they are human opinions embodied into a code), available at: <https://ampa.lifo.gr/koinotita/i-tyfli-empistosyni-stoys-algorithmoys-syntirei-tis-yparchoyses-diakriseis/>; Prodromou, A (2019), 'Η αβάσταχτη ελαφρότητα των αλγορίθμων – Οι αλγόριθμοι ετικετών εικόνας σίγουρα δεν είναι αντικειμενικοί όταν επεξεργάζονται εικόνες ανθρώπων' (The unbearable lightness of algorithms – The algorithms of picture tags are not objective when processing human pictures), *Kathimerini* (18 November 2019) available at: www.kathimerini.com.cy/gr/apopseis/prosopikotites-stin-k/i-abastaxti-elafrotita-ton-algorithmon; Fortune Greece (2020) 'Η τεχνητή νοημοσύνη και ο κίνδυνος των ψηφιακών διακρίσεων – Το ολλανδικό παράδειγμα μιας μεθόδου η οποία χρειάζεται ιδιαίτερη προσοχή' (Artificial Intelligence and the risk of digital discrimination – The Dutch example of a method which needs special attention), (16 February 2020) available at: www.fortunegreece.com/article/i-techniti-noimosini-ke-o-kindinos-ton-psifiakon-diakriseon/.

394 See <https://eimin.lrv.lt/lt/veiklos-sritys/skaitmenine-politika/dirbtinis-intelektas>.

395 <https://ncpe.gov.mt/en/Pages/Upcoming-events-and-Developments.aspx>.

396 Rathenau Instituut (2020), *Response of the Rathenau Instituut to the European Commission's White paper on Artificial Intelligence* (15 May 2020) available at: www.rathenau.nl/sites/default/files/2020-05/Response%20Rathenau%20Institute%20EC%20Whitepaper%20on%20AI%2015May2020.pdf.

for a conversation between the Government and the Parliament on further policy efforts that should be made.³⁹⁷ In **Poland**, public discussions on AI and risks of discrimination have been driven by civil society organisations, including the Panoptykon Foundation, the ePaństwo Foundation and the Foundation Centrum Cyfrowe. These CSOs have played a key role through the publication of reports addressing, among other things, the question of algorithmic discrimination.³⁹⁸ In **Sweden**, the Equality Ombudsman has conducted a mapping exercise and identified several typical risks of discrimination in automatised decision making.³⁹⁹ In addition, the Government commissioned the Swedish Social Insurance Inspectorate (Inspektionen för socialförsäkringen) to conduct a specific study on the impact of algorithmic profiling on gender equality at the Swedish Social Insurance Agency, which concluded that gender biases represent a risk but do not necessarily lead to violations of the Swedish Discrimination Act since there might be an objective and reasonable justification.⁴⁰⁰ In the **United Kingdom**, both the Government and Parliament have engaged in a growing debate on the regulation and control of artificial intelligence, including in relation to the problems of bias and discrimination.⁴⁰¹

Although these public discussions and debates at national level show considerable activity in the field, many of them primarily focus on the potential of digitalisation and the economic opportunities linked to the development of artificial intelligence, and the need for national countries to be at the forefront of such developments, both in the public and private sectors. For a number of countries (**Bulgaria, Croatia, Cyprus, Estonia, Hungary, Liechtenstein, Portugal, Romania, Slovenia**), national experts have observed that no explicit discussion of the discriminatory risks of AI has taken place yet, although a general debate about AI has emerged. For many other countries, discussions on AI and ethics or AI and human rights have taken place, even if they are often general in nature and do not specifically or extensively focus on the question of equality and discrimination. For instance, in **Austria**, public discussions on the risks of AI have, according to the national expert, mainly addressed issues of data protection and data security. Only recently did the problem of algorithmic discrimination come to the fore with a campaign led by the NGO 'Epicenter. Works' against the AMS algorithm discussed in section 3.1.2.1, which classifies unemployed workers depending on their long-term and short-term chances of labour market integration and accordingly decides on the allocation of training resources. The public response to this problem has

397 For more information, see Gerards, JH (forthcoming), *Country report: the Netherlands* (Report commissioned by the EU Fundamental Rights Agency as part of its project *Artificial Intelligence, Big Data and Fundamental Rights* (<https://fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights>) (forthcoming, upon request on file with the author).

398 Niklas, J, Sztandar-Sztanderska, K and Szymielewicz, K (2015), *Profiling the unemployed in Poland: social and political implications of algorithmic decision making* (Eliza Jakubiak tr, report by Fundacja Panoptykon, Warsaw) available at: https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf; Škop, M and others (2019), *alGOVrithms. State of Play. Report on Algorithms Usage in Government-Citizens Relations in Czechia, Georgia, Hungary, Poland, Serbia and Slovakia* (Krzysztof Izdebski ed, Fundacja ePaństwo) available at: epf.org.pl/pl/wp-content/uploads/sites/2/2019/05/alGOVrithms-State-of-Play-Report.pdf; Mileszyk, N, Paszcza, B and Tarkowski, A (2019), *AlgoPolska. Zautomatyzowane podejmowanie decyzji w służbie społeczeństwu* (Algopolska. Automated decision-making at the service of society) (Report 07/2019 by Fundacja Centrum Cyfrowe and Klub Jagielloński, Kraków) available at: <https://centrumcyfrowe.pl/wp-content/uploads/sites/16/2019/12/Algopolska-raport.pdf>.

399 Since the Inspectorate did not investigate whether there was a justification, it could not give an answer as to whether the algorithmic profiling actually constituted discrimination. See *Profiling som urvalsmetod för riktade kontroller* (Profiling as a selection method for targeted controls), ISF Report 2018 (5), Stockholm, 14. See also Equality Ombudsman (2018) 'Kunskapsöversikt om användningen och utvecklingen av automatiserad databehandling med algoritmer (artificiell intelligens) och stordata och diskriminering eller risker för diskriminering' (Knowledge review of the use and development of automated data processing with algorithms (artificial intelligence) and big data and discrimination or risks of discrimination), Registration No LED 2018/387, Document 55, 13; *Promemoria om automatiserad databehandling med algoritmer och risker för diskriminering: inom rekrytering och kreditgivning* (Memorandum on automated data processing with algorithms and risks of discrimination: in recruitment and lending), Registration No 2018/387, Document 37 (4 October 2019) 8.

400 Inspektionen för socialförsäkringen (Swedish Social Insurance Inspectorate) (2018), *Profiling som urvalsmetod för riktade kontroller* (Profiling as a selection method for targeted controls), ISF Report 2018:5, Stockholm, available at: <https://inspsf.se/publikationer/rapporter/2018/2018-03-26-profiling-som-urvalsmetod-for-riktade-kontroller>; Swedish Social Insurance Inspectorate (2018), *Riskbaserade urvalsprofiler och likabehandling* (Risk based selection profiles and equal treatment), (Working Report 2018:1AR, Stockholm) available at: <https://inspsf.se/publikationer/rapporter/2018/2018-06-15-riskbaserade-urvalsprofiler-och-likabehandling>.

401 Government Office for Science (2016), *Artificial intelligence: opportunities and implications for the future of decision making* (Report GS/16/19, London) available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf.

taken the form of a research project funded by the Austrian Academy of Science, as further discussed in section 3.3.2.1.⁴⁰² In **Belgium**, the issue of algorithmic discrimination has remained largely unexplored in public discussions apart from some concerns expressed in relation to risks of discrimination in the insurance sector and in public unemployment agencies.⁴⁰³ In **Czechia**, limited public discussions have addressed the question of discrimination in the framework of the working group on human rights and new technologies under the Office of the Government and the national strategy for artificial intelligence.⁴⁰⁴ In **Finland**, discussions have concentrated on the use of AI by public authorities and include ethical issues,⁴⁰⁵ yet the topic of algorithmic discrimination has so far mainly been brought up by the Non-Discrimination Ombudsman's Office.⁴⁰⁶ In **Germany**, where the problem of algorithmic discrimination seems to have gained some attention, in particular when former Justice Minister Heiko Maas proposed a digital non-discrimination law in 2017,⁴⁰⁷ the German Women Lawyers' Association has reacted to the Commission's white paper by underlining that the regulation of algorithmic discrimination should take into account the risks and opportunities of gender categorisation in the use of algorithms (e.g. imposing a binary gender identification versus using gender categorisation as means to actively correct structural inequalities).⁴⁰⁸ In **Iceland**, the national expert has reported that debates on the impact of algorithms on gender equality have not come to the fore and that institutional interests lie in the related but distinct issue of the automation of the labour market.⁴⁰⁹ In **Ireland**, a general public consultation on the development of a national strategy on artificial intelligence took place in October 2019, and the strategy is likely to include

402 See www.oew.ac.at/en/ita/projects/ams-algorithm/.

403 The national expert notes a parliamentary question to the Federal Minister for Employment, the Economy and Consumers, with responsibility for combating poverty, equal opportunities and disabled persons, regarding the discriminatory risks linked to the use of algorithms in the insurance sector and the regulation measures adopted: see Depraetere, M, sp.a (2019), *Algorithmes dans le secteur des assurances* (Algorithms in the insurance sector), 6 November 2019, (Question et réponse écrite No. 0122, législature: 55). In response, the Minister referred to the ethical guidelines drafted by the High-Level Expert Group on Artificial Intelligence of the European Commission for the use of data and declared herself favourable to a harmonised European approach on the use of algorithms in the insurance sector, see www.lachambre.be/QRVA/pdf/55/55K0006.pdf, 337; Interview conducted by Belgian expert Nathalie Wuïame with B Miller (Head of the digital transformation unit), M Verasso and F Zibouh (diversity unit), I El Hamli (Case manager), Actiris online meeting (10 June 2020) and with G Vanhumbeeck (Director for Innovation), K Scheerlinck (AI team leader) and V Buekenhout (Data Protection Manager), VDAB, online Teams meeting (12 June 2020).

404 Working group on Human Rights and New Technologies (2019), *Regulatory Framework for Artificial Intelligence in the European Union* (Position paper on the EU white paper), available at: http://observatory.law.cas.cz/wp-content/uploads/2020/02/CZ%20nonpaper_%20EU%20regulatory%20framework%20for%20AI.pdf; Czechia (2019) *Národní strategie umělé inteligence v České republice* (National artificial intelligence strategy), 4 available at: www.vlada.cz/assets/evropske-zalezitosti/umela-inteligence/NAIS_kveten_2019.pdf. In his foreword, Andrej Babiš, Prime Minister of the Czech Republic, writes: 'We will focus on the protection of each person and consumer, their rights and privacy, especially the weakest. We will prevent discrimination, manipulation and misuse of AI, we will set rules for decision-making algorithms about people in everyday life.'

405 Koivisto, R and others (2019), *Tekoäly viranomaistoiminnassa – eettiset kysymykset ja yhteiskunnallinen hyväksyttävyy* (Artificial intelligence in authority use – ethical and societal acceptance issues) (Publications of the Government's analysis, assessment and research activities 14/2019, Prime Minister's Office, 1 February 2019) available at: <http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161345/14-2019-Tekoaly%20viranomaistoiminnassa.pdf?sequence=1&isAllowed=y>.

406 Valonen, T (2020), 'Tiesitkö, että tekoälyyn ja algoritmeihin liittyvää syrjintää valvoo yhdenvertaisuusvaltuutettu?' (Did you know that discrimination related to artificial intelligence and algorithms is monitored by the Non-Discrimination Ombud?) *Yhdenvertaisuusblogi* (non-discrimination blog), (29 April 2020).

407 See Fröhlich, W and Spiecker genannt Döhm, I (2018), 'Können Algorithmen diskriminieren?' (Can algorithms discriminate?) *Verfassungsblog* (26 December 2018) available at: verfassungsblog.de/koennen-algorithmen-diskriminieren/; 'On the positive side, discrimination is recognised as a problem of algorithm-based decisions' (translation by expert). See also Beuth, P (2017), 'Maas schlägt digitales Antidiskriminierungsgesetz vor' (Maas proposes a digital non-discrimination law) *Zeit Online* (3 July 2017) available at: www.zeit.de/digital/internet/2017-07/heiko-maas-algorithmen-regulierung-antidiskriminierungsgesetz, and criticism by NGOs, see Jaume-Palasi, L (2017), 'Diskriminierung hängt nicht vom Medium ab' (Discrimination does not depend on the medium) *AlgorithmWatch*, (3 July 2017) available at: <https://algorithmwatch.org/diskriminierung-haengt-nicht-vom-medium-ab/>.

408 German Women Lawyers' Association (2020), 'Stellungnahme zu dem Weißbuch der EU-Kommission „Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“' (Statement on the EU Commission white paper on artificial intelligence – a European concept for excellence and trust) (Berlin, 9 June 2020) available at: www.djb.de/fileadmin/user_upload/presse/stellungnahmen/st20-20_WeissbuchKI_EU.pdf.

409 Þorsteinsson, HF and others (2019), *Iceland and the fourth industrial revolution* (Committee on the Fourth Industrial Revolution, Government of Iceland, Prime Minister's Office, October 2019) available at: www.government.is/lisalib/getfile.aspx?itemid=3fe9f512-ef4-11e9-944e-005056bc530c.

issues around ethics, inclusion and diversity in artificial intelligence among other things.⁴¹⁰ In **Italy**, the national expert mentions an emerging public debate in the wake of the two court cases mentioned in section 3.1.3.1 (on delivery services and work placements for teachers) and international initiatives,⁴¹¹ including the consultation process linked to the European Commission's white paper.⁴¹² In **Latvia**, discussions on the impact of algorithms on equality are emerging on the basis of the guidelines published by the European High-Level Expert Group on AI⁴¹³ and the European Commission's 2018 Coordinated Plan on the Development of Artificial Intelligence Made in Europe,⁴¹⁴ as well as on conclusions adopted by the EU Council⁴¹⁵ and a policy document addressing among other things issues of algorithmic discrimination that was adopted by the Cabinet of Ministers in 2020.⁴¹⁶ In **Luxembourg**, the expert indicates a general lack of awareness of the issue despite political declarations on 'artificial intelligence for the benefit of all'.⁴¹⁷ However, the Luxembourg Government was asked by means of a Parliamentary motion to monitor the impact of new technologies in order to guarantee non-discrimination⁴¹⁸ at the end of a 2019 Parliamentary debate on the 'Digital transformation of Luxembourg'.⁴¹⁹ In **Norway**, awareness of the topic of AI and equality is increasing,⁴²⁰ but knowledge of AI and algorithmic discrimination reportedly remains low, especially among legal practitioners and members of the judiciary. In **Slovakia**, there appear to be very few discussions on algorithmic discrimination in the public sphere.⁴²¹ Finally, the national expert

- 410 Department of Business, Enterprise and Innovation (2019), 'Four in ten Irish companies currently use Artificial Intelligence (AI)' (8 April 2019) available at: <https://dbe.gov.ie/en/News-And-Events/Department-News/2019/April/09042019.html>.
- 411 These include e.g. the Council of Europe Decalogue on Algorithms and Human Rights – The impact of automated decisions and possibilities for their discipline, see 'Gli algoritmi possono anche discriminare' (Algorithms can also discriminate) ANSA (17 July 2017) available at: www.ansa.it/sito/notizie/tecnologia/software_app/2017/07/16/algoritmi-possono-anche-discriminare_9222ccad-ccb0-4863-ad64-f5aec1c3ffb4.html; Morelli, C (2018), 'Algoritmi e diritti umani: si rischia la collisione (immaginate a danno di chi?)' (Algorithms and human rights: the risk of collision (imagine to whose detriment?)), *Altalex*, (9 July 2018), available at: <https://www.altalex.com/documents/news/2018/07/09/algoritmi-e-impatto-sui-diritti-umani>.
- 412 See for instance the specialised website www.agendadigitale.eu, but it seems that discussions focus more on data protection issues than on discrimination: 'While the participants at the debate on privacy and the same laws on the subject focus on the protection of people, and therefore of identifying data, the risks of the processing of anonymised and aggregated information that, instead, appear to be extremely important, and of discrimination on groups and minorities are not taken into consideration' (translation by expert); see Saetta, B (2018), 'Algoritmi, intelligenza artificiale, profilazione dei dati: cosa rischiamo davvero come cittadini?' (Algorithms, artificial intelligence, data profiling: what do we really risk as citizens?) *Valigiablu*, (8 January 2018) available at: www.valigiablu.it/algoritmi-dati-rischi/; see also Panorama (2019) 'Internet: gli algoritmi e le discriminazioni' (Internet: algorithms and discrimination) (11 April 2019) available at: www.panorama.it/internet-gli-algoritmi-le-discriminazioni; Numerico, T (2019), 'Algoritmi, strumenti opachi' (Algorithms, opaque instruments), *Il Manifesto* (30 June 2019) available at: ilmanifesto.it/algoritmi-strumenti-opachi/. Examples of reactions to the EC white paper at the national level include Ministry of Economic Development (2019), *Strategia Nazionale per l'Intelligenza Artificiale* (National Strategy for Artificial Intelligence) available at: www.mise.gov.it/images/stories/documenti/Strategia-Nazionale-Intelligenza-Artificiale-Bozza-Consultazione.pdf, which only marginally mentions the risk of discrimination, and AGID (Agenzia per l'Italia Digitale) (2018), *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino* (White paper on AI at the service of the citizen) (Version 1.0) available at: ia.italia.it/assets/librobianco.pdf, which tackles the issue more extensively.
- 413 European Commission (2018) 'Have your say: European expert group seeks feedback on draft ethics guidelines for trustworthy artificial intelligence' (10 December 2018) available at: <https://ec.europa.eu/digital-single-market/en/news/have-your-say-european-expert-group-seeks-feedback-draft-ethics-guidelines-trustworthy>.
- 414 European Commission (2018), 'Coordinated Plan on Artificial Intelligence' (7 December 2018) available at: <https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence>.
- 415 Council of the European Union, 'Conclusions on the coordinated plan on artificial intelligence' (11 February 2019) (6177/19) available at: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/en/pdf>.
- 416 See Press release by the Minister of Environmental Protection and Regional Development, available in Latvian at: www.mk.gov.lv/lv/aktualitates/jpuce-maksliga-intelektas-risinajumi-jaizmanto-tautsaimniecibas-produktivitates-celsanai; VARAM Informatīvais ziņojums *Par mākslīgā intelekta risinājumu-jaizmanto-tautsaimniecības-produktivitātes-celsanai* (Informative report of the Ministry of Environmental Protection and Regional Development Ministry on development of artificial intellect) (Conclusions, Section 6) 51, available at: www.varam.gov.lv/lv/jauns-informativais-zinojums-par-maksliga-intelektas-risinajumu-attistibu.
- 417 Luxembourg Ministry of Digitalisation (2019) 'Artificial intelligence for the benefit of all', Interview with Xavier Bettel in 'Happen' (20 December 2019) available at: <https://digital.gouvernement.lu/fr/actualites/gouvernement%2Ben%2Bactualites%2Binterviews%2B2019%2B12-decembre%2B20-bettel-happen.html>.
- 418 <https://chd.lu/wps/wcm/connect/public/f0475119-2fe0-4801-bf04-efa2bcebc323/motionberger.pdf?MOD=AJPERES&ContentCache=NONE&CACHE=NONE&CVID=mKV2Ro-&CVID=mKV2Ro-&CVID=mKV2Ro-&CVID=mKV2Ro-&CVID=mKV2Ro->>.
- 419 www.chd.lu/ArchivePlayer/video/2580/sequence/126676.html.
- 420 Teknologirådet (2018), *Kunstig Intelligens – muligheter, utfordringer og en plan for Norge* (AI, possibilities, challenges and a plan for Norway) (Oslo, September 2018) available at: <http://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf> 52-53, 74.
- 421 The action plan adopted by the Government only makes a general mention of algorithmic discrimination, see Slovakia (2019) *Action plan for the digital transformation of Slovakia for 2019-2022* (Government Resolution No. 337/2019, 3 July 2019) available in English at: www.vicepremier.gov.sk/wp-content/uploads/2019/10/AP-DT-English-Version-FINAL.pdf.

for **Spain** has signalled that although interest in the ethical issues linked to the use of AI is growing, the attention dedicated to the impact of algorithms on gender equality and the question of algorithmic discrimination is still limited. Public discussions on the topic have largely remained theoretical or based on cases in foreign jurisdictions or have focused on issues related to data protection and privacy.⁴²²

All in all, although the situation varies across Europe, the national reports show that public and political awareness of discrimination issues linked to algorithms is rather limited and the topic is only just emerging in the public space of the 31 countries covered by this report. In addition, in a majority of countries, these discussions are limited to a rather abstract level: algorithmic bias is acknowledged as a general risk and ethical concerns for human rights in the age of AI are recognised as transversal challenges, but often at the margin of much broader discussions that are usually focused on economic growth and the development of a national AI strategy. Moreover, it can be observed that these public discussions are strongly influenced by US-centred literature, media coverage and policy debates, which is evident in the examples used to illustrate risks of algorithmic discrimination and in the focus of national public discourse. For example, the example of race discrimination in policing and the American COMPAS algorithm used in this respect is cited regularly, whereas European examples are mentioned much less often. As further analysed in Chapter 4, this influence could prove problematic because US discrimination law and core concepts are different from those of EU law, as are usages of algorithms, notably in relation to the public/private divide. More distinctive public debates are gradually emerging, as the Austrian discussion of the AMS algorithm has shown,⁴²³ but this trend is rather recent.

3.3.1.2 Interactions between data protection law and gender equality and non-discrimination law

Beyond generic discussions, there seems to be limited awareness in European countries' public space of specific issues such as those relating to the interaction between data protection law and gender equality and non-discrimination law. While issues of data protection and privacy are often flagged as central self-standing issues in national debates, a majority of national experts report that the interaction between the two bodies of norms – data protection law and equality law – have not been explored in the public domain

422 Media articles remain generalist and focused on international examples and cases. See e.g. Carballar, O (2020), 'Los algoritmos pueden intensificar la desigualdad' (Algorithms may render inequality stronger) *La Marea* (9 February 2020) available at: www.lamarea.com/2020/02/09/big-data-privacidad-desigualdad/; Del Castillo, C (2019), 'Carlos Preciado: "Los algoritmos a veces dan una falsa impresión de objetividad que aplasta los derechos de las personas"' (Carlos Preciado: 'Sometimes algorithms give a false impression of objectivity that quashes people's rights') *El Diario* (9 July 2019) available at: www.eldiario.es/tecnologia/algoritmo-maquinas-impresion-objetividad-derechos_0_918658323.html; Pascual, MG (2019), 'Quién vigila que los algoritmos no sean racistas o sexistas?' (Who monitors that algorithms are not racist or sexist?) *Revista Retina* (17 March 2019) available at: https://retina.elpais.com/retina/2019/03/14/tendencias/1552564034_268678.html; Romero, P (2019), 'Lo que el algoritmo esconde: matemáticas, sesgos y al final, decisiones humanas' (What the algorithm hides: mathematics, bias and finally, human decisions) *Público* (8 February 2019) available at: www.publico.es/ciencias/inteligencia-artificial-algoritmo-esconde-matematicas-sesgos-final-decisiones-humanas.html; Quincoces Riesco, A (2018), 'El lado oscuro de los algoritmos' (The dark side of algorithms), *La Vanguardia* (11 June 2018) available at: www.lavanguardia.com/tecnologia/20180611/4518476004/el-lado-oscuro-de-los-algoritmos.html; Romero, P (2018) 'Manual para evitar "tecnologías machistas"' (Textbook to avoid 'male-chauvinist technologies') *Público* (29 March 2018) available at: www.publico.es/sociedad/cibersociedad-manual-evitar-tecnologias-machistas.html; Duarte, F (2018), '5 algoritmos que ya están tomando decisiones sobre tu vida y que quizás tú no sabías' (5 algorithms that are already taking decisions on your life and you may not know it) *BBC* (3 February 2018), available at: www.bbc.com/mundo/noticias-42916502; Aishow (2017), 'El Gobierno crea un Grupo de Sabios para elaborar un libro blanco sobre Inteligencia Artificial y Big Data' (The Government sets up a group of experts to elaborate a white paper on artificial intelligence and big data) (15 November 2017) available at: <https://the-aishow.com/news/gobierno-crea-grupo-sabios-elaborar-libro-blanco-inteligencia-artificial-big-data/>. Reports that address the question of automated decision making in Spain do not extensively analyse the societal impact of AI, see e.g. Spielkamp, M (2019), *Automating Society. Taking Stock of Automated Decision-Making in the EU* (Report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations, 1st edn, January 2019) available at: https://algorithmwatch.org/wp-content/uploads/2019/01/Automating_Society_Report_2019.pdf; *Consulta pública sobre la estrategia digital para una España inteligente* (2017) (Public consultation on the digital strategy for a smart Spain) available at: <https://avancedigital.gob.es/es-ES/Participacion/Documents/estrategia-digital-espana-inteligente.pdf>. One of the rare instances in which concern regarding the perpetuation of gender discrimination by algorithms can be found in Mateos Sillero, S and Gómez Hernández, C (2020), *Libro Blanco de las mujeres en el ámbito tecnológico* (White paper on women in technology), Secretary of State for Digital Advancement, available at: www.mineco.gob.es/stfls/mineco/ministerio/ficheros/libreria/LibroBlancoFINAL.pdf.

423 See further section 3.3.1.1.

(Austria, Belgium, Bulgaria, Croatia, Cyprus, Estonia, Hungary, Ireland, Latvia, Liechtenstein, Lithuania, Portugal, Romania, Slovakia, Slovenia).

In countries where discussions have tackled the interaction between data protection and discrimination, there is a clear consensus on the relevance of the General Data Protection Regulation (GDPR) and in particular its Articles 5 (on the treatment of personal data), 13-15 (on rights of information and access to personal data) and 22 (the right not to be subject to an individual decision based solely on automated processing, including profiling). In **France**, the equality body (the Defender of Rights) and the data protection authority (*Commission Nationale Informatique et Libertés*) have recently published a joint report that clearly addresses the intersection between these two regulation bodies, underlining the necessity of cooperation and coordination between corresponding agencies and highlighting the limits of both data protection and equality law.⁴²⁴ In the **Netherlands**, the interaction between data protection and equality law has gained visibility, notably with respect to discrimination on the ground of race or ethnic origin. A debate has taken place on the use of information on data related to such grounds (e.g. having a second nationality) by tax authorities, municipalities and other public organisations in risk analyses (especially for fraud detection purposes).⁴²⁵ In **Sweden**, the Equality Ombudsman has questioned the extent to which the GDPR may be used by an individual to gain access to information to assess (and prove) the occurrence of algorithmic discrimination.⁴²⁶ In other countries there are more generic discussions on the interaction between data protection and equality law, for instance **Finland**, where the Office of the Data Protection Ombudsman has declared that impact assessments of the outcome of data processing involving natural persons can serve as a tool for detecting algorithmic discrimination,⁴²⁷ or **Poland**, where civil society has underlined the importance of data protection law in combating algorithmic discrimination and other breaches of fundamental rights.⁴²⁸

3.3.1.3 Issues of liability and responsibility for algorithmic discrimination

Another specific issue reported by national experts relates to questions of responsibility and liability for algorithmic discrimination. A majority of countries have witnessed public debates on generic questions of liability, including on the legal personhood of AI and possible forms of liabilities for legal breaches arising from the use of algorithms. However, in the vast majority of countries, these discussions have reportedly not been (extensively) linked to issues of algorithmic discrimination in public discussion (**Austria, Belgium,**

424 Defender of Rights and CNIL (2020), *Algorithms: prévenir l'automatisation des discriminations* (Algorithms: preventing automated discrimination) joint report, available at: www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-en-num-16.07.20.pdf.

425 There have been many publications in the media about this matter. See e.g. Kleinnijenhuis, J (2020), 'Belastingdienst erkent: toch sprake van etnisch profileren' (The Tax and Customs Administration acknowledges: ethnic profiling is happening) *Trouw* (10 May 2020) available at: www.trouw.nl/nieuws/belastingdienst-erkent-toch-sprake-van-etnisch-profileren~b91d1a45/; see also the report by the Advisory Committee on the Implementation of Benefits (2019), *Omzien in verwondering. Eindadvies* (Looking back in amazement. Final advice), part 2, (The Hague, 12 March 2020), annex No. 926526 to *Kamerstukken II* (Parliamentary Papers) 2019/20, 31066, No. 608 available at: www.rijksoverheid.nl/documenten/kamerstukken/2020/03/12/omzien-in-verwondering-eindadvies-adviescommissie-uitvoering-toeslagen.

426 Equality Ombudsman (2018), *Kunskapsöversikt om användningen och utvecklingen av automatiserad databehandling med algoritmer (artificiell intelligens) och stordata och diskriminering eller risker för diskriminering* (Knowledge review of the use and development of automated data processing with algorithms (artificial intelligence) and big data and discrimination or risks of discrimination) (Registration No LED 2018/387, Document 55) and Equality Ombudsman (2019), *Promemoria om automatiserad databehandling med algoritmer och risker för diskriminering: inom rekrytering och kreditgivning* (Memorandum on automated data processing with algorithms and risks of discrimination: in recruitment and lending) (Registration No. 2018/387, Document 37, 4 October 2019).

427 The Data Protection Ombudsman's Office's website provides guidelines on use of IT and profiling. The guidelines stress that any IT use that involves profiling must take certain minimum precautions, including a follow-up of the algorithms and data used in order to make sure that decision-making procedures do not involve discrimination; see Data Ombudsman's Office (2019) *Automaattinen päätöksenteko ja profilointi* (Automated decision making and profiling), available at: <https://tietosuojaja.fi/automaattinen-paatoksenteko-profilointi>.

428 See e.g. Panoptikon Foundation: https://panoptikon.org/sites/default/files/leadimage-biblioteka/panoptikon_profiling_report_final.pdf, 39; Centrum Cyfrowe Foundation: <https://centrumcyfrowe.pl/wp-content/uploads/sites/16/2019/12/Algopolska-raport.pdf> 52, ePaństwo Foundation: <https://epf.org.pl/pl/2019/10/28/manifest-epanstwa-postulaty-fundacji-epanstwo-na-kadencje-sejmu-2019-2023/>.

Bulgaria, Cyprus, Czechia, Denmark, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Portugal, Romania, Slovakia, Slovenia).

In some countries, issues of liability and responsibility for algorithmic discrimination have been the subject of some degree of public discussion. In **Croatia**, for instance, limited attention has been paid to the issue of liability in the context of online harassment and discrimination in the wake of an announcement made in 2018 regarding the preparation of a draft act on the prevention of inappropriate online behaviour, although no bill has yet been introduced.⁴²⁹ In **France**, various organisations have publicly addressed these issues, suggesting that shared liability should be carried by those who develop and those who use discriminatory algorithms, particularly in light of the increase of risks of discrimination linked to deep-learning algorithms.⁴³⁰ However, several organisations underline the importance of prevention strategies for algorithmic discrimination.⁴³¹ In contrast to the French position, the guidelines of the **Spanish** Agency for Data Protection establish that a decision maker relying on AI cannot avoid responsibility by arguing that it has insufficient access to information or technical knowledge regarding the functioning of the system, which in turn means that the responsibility cannot be allocated to the developer of an algorithm and even less to the AI system itself, but rather that the end user is responsible for testing and auditing the system for compliance with relevant regulations.⁴³² Other arguments, for instance in **Norway**, are that '[i]n order for a person to be able to take responsibility, in many cases it will be necessary for that person to understand how the algorithm makes recommendations. In those situations where an explanation is important but not present, it may therefore be necessary to choose less precise algorithms, which can however provide explanations'.⁴³³

National discussions also concern the question of whether current regimes of liability and sanctions are adequate to tackle algorithmic discrimination. In **Poland**, the idea has been put forward that the current regime of liability does not meet the new needs related to the use of AI.⁴³⁴ In particular, it has been argued

429 Prior evaluation of effects of legislation was conducted, and adoption of the act was planned in 2019. See e-consultations, available at: <https://esavjetovanja.gov.hr/ECon/MainScreen?entityId=9137>. The main objective of the act was stated to be the adoption of measures to prevent, detect and prohibit access to illegal online content and to increase transparency and protection of fundamental rights in social networks. There is an expectation for EU guidance on the subject, see People's Ombudsman (2020), *Izvešće pučke pravobraniteljice za 2019* (Annual Report for 2019) (March 2020, Zagreb) available at: www.ombudsman.hr/wp-content/uploads/2020/03/Izve%C5%A1%C4%87e-pu%C4%8Dke-pravobraniteljice-za-2019.pdf, 126-127.

430 www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-num-05.06.20.pdf, 9.

431 See Institut Montaigne (2020), *Algorithmes : contrôle des biais S.V.P.* (Algorithms: Bias control, please) (Report, March 2020) available at: www.institutmontaigne.org/ressources/pdfs/publications/algorithmes-contrôle-des-biais-svp.pdf; President Macron promised to follow the proposals of the Villani Report in a speech at the Collège de France in 2018: 'Discours du Président de la République sur l'intelligence artificielle' (Presidential speech on artificial intelligence) (29 March 2018) available at: www.elysee.fr/emmanuel-macron/2018/03/29/discours-du-president-de-la-republique-sur-lintelligence-artificielle. See also more recently 'Clôture du Global Forum on Artificial Intelligence for Humanity par le Président Emmanuel Macron' (Closing of the Global Forum on Artificial Intelligence for Humanity by President Emmanuel Macron) (30 October 2019) available at: www.elysee.fr/emmanuel-macron/2019/10/30/cloture-du-global-forum-on-artificial-intelligence-for-humanity-par-le-president-emmanuel-macron; Villani, C (2018), *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne* (Giving meaning to artificial intelligence. For a national and European strategy) (Report for the Prime Minister, September 2018) available at: www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf; Defender of Rights and CNIL (National Data Protection Authority of France) (2020), *Algorithmes: prévenir l'automatisation des discriminations* (Algorithms: preventing automated discrimination) joint report, available at: www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-en-num-16.07.20.pdf.

432 Spanish Agency on Data Protection (2020), *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción* (Adaptation to the GDPR of treatments that incorporate artificial intelligence. An introduction) (February 2020) www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf.

433 Teknologirådet (2018), *Kunstig Intelligens – muligheter, utfordringer og en plan for Norge* (AI, possibilities, challenges and a plan for Norway) (Oslo, September 2018) available at: <https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf>, 52-53, 74.

434 See *Polityka Rozwoju Sztucznej, Inteligencji w Polsce na lata 2019 – 2027* (Policy for the Development of Artificial Intelligence in Poland for 2019-2027) available at: www.gov.pl/attachment/0aa51cd5-b934-4bcb-8660-bfecb20ea2a9, 102-3. The policy document emphasised that the provisions of private law on liability for damages are not adapted to the challenges posed by AI, and it was pointed out that the resolving of the emerging problems could be attempted on a micro scale, trying to find temporary solutions beforehand. According to the policy document, ongoing activities should consist of an appropriate adjustment of the regulations on liability for dangerous products. Nevertheless, in the long term entirely new rules of civil liability for algorithms should be developed. Finally, the document emphasised that it would be optimal to establish these rules by way of international consensus.

that the liability regime for dangerous products should be extended to cover different types of digital content, and that this reform should be conducted at the EU level based on Directive 85/374.⁴³⁵ Such an ad hoc solution would assimilate the discriminatory nature of an algorithmic design to a ‘defective’ product. In **Sweden**, the Equality Ombudsman has highlighted the lack of adequate effective sanctions for group-based discrimination as a particular problem in light of the rapid development of automated decision making: ‘In such systems, [...] algorithms can [cause] discrimination [against] a very large number of individuals without them or anyone using the system knowing it. As in the case of normal discriminatory rules, there must be an opportunity to effectively counteract algorithms that risk leading to discrimination at the group level’.⁴³⁶

3.3.2 *Scientific discussions on the impact of algorithms on gender equality and non-discrimination*

3.3.2.1 General legal scholarship on the discriminatory impact of AI

Scholarship on the discriminatory impact of AI appears to be emerging in all countries covered in this report. When scientists and academics, in particular legal researchers, address the topic, the focus seems to overwhelmingly remain on the US context or international regulatory or ethical initiatives (for example at the level of the Council of Europe) rather than on national contexts. In addition, experts report that a majority of national scholarship on AI concerns issues other than discrimination, such as data protection, privacy, public administration, the legal personhood of AI, etc. The scientific analysis of the specific impact of AI on gender equality and non-discrimination law at national level – including questions of whether national law still fits the bill or if specific risks arise in relation to the use of algorithmic decision making at national level – seems to remain limited. Empirical data on, and scientific analysis of, the impact of algorithms on discrimination in national contexts remain scarce.

For **Bulgaria, Croatia, Cyprus, Czechia, Estonia, Hungary, Latvia, Liechtenstein, Malta, Romania, Slovakia** and **Slovenia**, national experts have reported no extensive scientific discussion of the impact of algorithms on gender equality and non-discrimination law. They nevertheless expect that this will change in the near future given that an increasing number of universities are establishing centres for the study of artificial intelligence (for example in **Croatia**⁴³⁷ and **Estonia**),⁴³⁸ consultancy and law firms are offering specialised services⁴³⁹ and the use of algorithms in national markets is progressively growing.⁴⁴⁰

Emerging research on this topic has been reported in a number of countries, where the number of scientific events, research projects as well as PhD and masters theses on algorithmic discrimination has increased over recent years. In **Austria**, research into the discriminatory impact of the AMS algorithm, the Austrian public employment service’s intended profiling system,⁴⁴¹ has been conducted by the Centre for

435 Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [1985] OJ L210/29.

436 Diskriminerings ombudsmannen (DO) (2019), ‘Vi behöver mer effektiva och avskräckande sanktioner’ (We need more effective and deterrent sanctions) (Press release, 18 December 2019) available at: www.do.se/om-do/pressrum/aktuellt/aktuellt-under-2019/vi-behoover-mer-effektiva-och-avskrackande-sanktioner/; see also the DO’s submission to the Government: Diskriminerings ombudsmannen (DO) (2019), ‘Om behovet av effektiva och avskräckande sanktioner vid vissa överträdelser av diskrimineringslagen (2008:567) med mera’ (On the need for effective and deterrent sanctions concerning certain violations of the Discrimination Act (2008:567) etc.) (16 December 2019) available at: www.do.se/om-do/vad-gor-do/skrivelser-till-regeringen/om-behovet-av-effektiva-och-avskrackande-sanktioner-vid-vissa-overtradelser-av-diskrimineringslagen-2008567-med-mera/.

437 See e.g. the Centre for Artificial Intelligence of the Zagreb Faculty of Electrical Engineering and Computing, <https://cai.fer.hr/en/cai> and the Centre for Artificial Intelligence and Cyber Security of the University of Rijeka, <https://airi.uniri.hr>.

438 See e.g. the IT Law Lab <https://oigus.ut.ee/en/admissions/challenges-field-it-law>.

439 See e.g. in Croatia Parser Compliance <https://parser.hr/pravo-i-algoritmi/>; <https://parser.hr/en/regulating-algorithms/>.

440 Ságvári, B (2017), ‘Diskrimináció, átláthatóság és ellenőrizhetőség. Bevezetés az algoritmusetikába’ (Discrimination, transparency and control: An introduction of algorithm ethics) 103(3) *Replika* available at: http://replika.hu/system/files/archivum/replika_103-04_sagvari.pdf, 61, 77.

441 For a more detailed description, see section 3.1.2.1.

Informatics and Society at the TU Wien and funded by the Austrian Academy of Sciences.⁴⁴² In **Belgium**, the Knowledge Centre at the University of Leuven recently organised a webinar on data and recruitment, which dealt with the question of algorithmic bias in HR and recruitment processes.⁴⁴³ The Open University of **Cyprus** has recently established a specialised Centre for Algorithmic Transparency, which reflects a growing interest in understanding the phenomenon of algorithmic discrimination. In **Finland**, too, some research exists,⁴⁴⁴ although legal scholarship on algorithmic discrimination remains scarce. For example, a report commissioned by the Prime Minister's Office concludes that Finnish legal scholarship does not extensively address the legal risks arising from the increasing use of AI.⁴⁴⁵ **French** legal scholarship has addressed questions of human rights and AI, for example in relation to the ethics of predictive justice in civil law,⁴⁴⁶ although the specific impact of algorithms on French gender equality and non-discrimination law has not yet been extensively covered. Furthermore, some interdisciplinary studies have been conducted⁴⁴⁷ and a symposium on the topic of algorithms and discrimination was held in Paris in December 2018, cosponsored by the National Conservatory of Arts and Crafts (Conservatoire national des arts et métiers) and the French think tank, Trans Europe Experts, which gave rise to an edited volume on the topic.⁴⁴⁸ Future developments are expected as conferences and scientific events on the topic of algorithmic discrimination are increasingly taking place in the country's scientific institutions.⁴⁴⁹ In **Greece**, scientific research to date has mainly addressed questions of data protection and privacy, although risks of discrimination have been acknowledged in relation to various fields of application,⁴⁵⁰ for

442 www.oeaw.ac.at/en/ita/projects/ams-algorithm/.

443 Knowledge Centre Data & Society (2020), 'Data-Date #2: AI en rekrutering' (Data-Date #2: AI and recruitment) (29 April 2020), available at: <https://data-en-maatschappij.ai/en/news/data-date-2-ai-en-rekrutering>.

444 Nurmilaakso, T (2017), 'Tekoäly valitsee jo Suomessakin parhaat työntekijät – Tutkijat: "Lisää syrjintää, seksismä ja rasismia"' (Artificial intelligence in Finland already selects the best employees – Researchers: "Increasing discrimination, sexism and racism") Yle Finnish Broadcasting Company, (7 October 2017) available at: <https://yle.fi/aihe/artikkeli/2017/10/07/tekoaly-valitsee-jo-suomessakin-parhaat-tyonhakijat-tutkijat-lisaa-syrjintaa>.

445 See for example Koulu, R and others (2019), *Algoritmi päätöksentekijänä? Tekoälyn hyödyntämisen mahdollisuudet ja haasteet kansallisessa päätöksenteossa* (Algorithm as a decision-maker? The possibilities and challenges of artificial intelligence in the national regulatory environment) (Prime Minister's Office, 2019:44 Helsinki, 25.7.2019) available at: http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161700/TEAS_44_19_Algoritmi%20paatoksenteekijana.pdf; Koulu, AR (2018), 'Digitalisaatio ja algoritmit – oikeustiede hukassa?' (Digitalisation and algorithms – Legal science getting lost?) 116(7-8) *Lakimies* 28, 840.

446 See e.g. Dondero B (2017), 'Justice predictive: la fin de l'aléa judiciaire?' *Recueil Dalloz* 532; Sentis T (2018) *Les enjeux éthiques de la Justice Predictive. Livre* (Ethical Issues in Predictive Justice) (White paper by SciencesPo École de Droit and Predictice) available at: www.sciencespo.fr/ecole-de-droit/en/actualites/la-justice-predictive-peut-elle-etre-ethique.html; Amrani-Mekki, S (2018), 'Justice predictive et accès au juge' (Predictive justice and access to the judge) in Amrani-Mekki, S and others, *La justice prédictive. Actes du colloque du 12 février 2018 organisé par l'Ordre des avocats au Conseil d'Etat et à la Cour de cassation à l'occasion de son bicentenaire, en partenariat avec l'Université Paris-Dauphine PSL* (Predictive justice. Proceedings of the symposium of February 12, 2018 organized by the Bar Association at the Council of State and the Court of Cassation on the occasion of its bicentennial, in partnership with the University Paris-Dauphine PSL) (Dalloz); Cadet, L (2017), *L'Open data des décisions de justice. Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice* (Open data of court decisions. Mission of study and prefiguration on the opening to the public of court decisions) (Rapport à Madame la garde des Sceaux, ministre de la Justice, November 2017) available at: www.justice.gouv.fr/publication/open_data_rapport.pdf.

447 See e.g. Bertail, P and others (2019), *Algorithmes: biais, discrimination et équité* (Algorithms: bias, discrimination and equity) (Telecom ParisTech and Fondation Abeona, February 2019) available at: www.telecom-paris.fr/wp-content/uploads/2019/02/Algorithmes-Biais-discrimination-equite.pdf.

448 See Mercat-Bruns, M (2020), *Nouveaux modes de détection et de prévention de la discrimination et accès au droit. Action de groupe et discrimination systémique. Algorithmes et préjugés. Réseaux sociaux et harcèlement* (New ways of detecting and preventing discrimination and access to the law. Class action and systemic discrimination. Algorithms and prejudices. Social networks and harassment) (1st edn, Société de législation comparée ed), *Collection Trans Europe Experts*, Vol 14) available at: www.lgdj.fr/nouveaux-modes-de-detection-et-de-prevention-de-la-discrimination-et-acces-au-droit-9782365170925.html.

449 'Les Convergences du Droit et du Numérique' (The Convergence of Law and the Digital World) (Université de Bordeaux, 16/17 March 2020, postponed) available at: https://cdn.u-bordeaux.fr/files/convergences/programme_ateliers_2020_v0.4_web.pdf; 'Workshop on algorithms and discrimination' Saclay (Region of interdisciplinary applied science institutes of research and Ecole Normale Supérieure Cachan, see Law Professor F Marty (University of Nice) on Law and Economics, Bias and algorithms, 12 December 2017), available at: <http://msh-paris-saclay.fr/wp-content/uploads/2017/11/Programme-12-December-Algorithm-bias-and-discrimination.pdf>.

450 Symeonidis, I (2019), 'Τεχνητή Νομοσύνη και Δικαιοσύνη' (Artificial Intelligence and Justice), Εφημερίδα Διοικητικού Δικαίου (ΕΔΔ) (*Administrative Law Journal*) 462; Tassis, S (2018), 'Η εποχή της τεχνητής νομοσύνης' (The era of artificial intelligence), Δίκαιο Μέσων Ενημέρωσης και Επικοινωνίας (ΔΙΜΕΕ) 4 (*Media and Communication Law*) 484-494; Apostolidis, C (2019), 'A legal approach to Algorithm-driven process and Artificial Intelligence' (LLM thesis, International Hellenic University) available at: https://repository.ihu.edu.gr/xmlui/bitstream/handle/11544/29440/c.apostolidis_ilm_16.04.2019.pdf?sequence=1.

example in relation to work,⁴⁵¹ criminal sentencing,⁴⁵² and facial recognition.⁴⁵³ The challenges such risks of discrimination pose for judges and legislators have also been recognised⁴⁵⁴ and some authors have argued that public interest should override principles of intellectual property and trade secrets when it comes to accessing algorithmic systems for the purposes of human rights protection.⁴⁵⁵ In **Iceland**, the incorporation of human biases into AI systems and the lack of diversity in the IT workforce have gained some attention, for example on the occasion of a scientific event entitled 'Are there inherent prejudices in artificial intelligence' held in October 2019 at Reykjavík University and sponsored by the NGO VERTOnet, which was established in 2018 to guard the interests of women in the IT sector.⁴⁵⁶ In **Ireland**, the issue of AI and gender bias has attracted some academic attention⁴⁵⁷ and some major law firms have published reports touching on issues of algorithmic discrimination in recruitment processes.⁴⁵⁸ In **Italy**, although legal scholarship has not yet extensively dealt with the question, change is expected as scientific projects have recently been funded to look into the problem of algorithmic discrimination. For instance, the project 'NoBIAS – Artificial Intelligence without Bias' conducted by Professor S. Ruggieri at Pisa University is expected to train 15 researchers through a multidisciplinary approach combining computer science, data science, machine learning and legal and social sciences.⁴⁵⁹ In some countries, legal scholarship is emerging on this topic, as shown by the organisation of scientific events and ongoing PhD projects, for example in **Lithuania**.⁴⁶⁰ In **Sweden**, the impact of algorithms on gender equality and non-discrimination has been addressed in graduate theses, which have partly analysed the issue in the context of Swedish

-
- 451 Douka, V (2020), 'Η λήψη «πλήρως αυτοματοποιημένων αποφάσεων» από τον εργοδότη: Το άρθρο 22 του Γενικού Κανονισμού Προσωπικών Δεδομένων (Κανονισμός 2016/679- Γ.Κ.Π.Δ.)' (The employer's decisions based solely on automated processing: Article 22 of the General Data Protection Regulation (Regulation 2016/679 – GDPR)) 3 Επιθεώρησης Εργατικού Δικαίου (ΕΕργΔ) (*Labour Law Review*) 369; Travlos-Tzanetatos, D (2019), Το εργατικό δίκαιο στην τέταρτη βιομηχανική επανάσταση (*Labour Law in the Fourth Industrial Revolution*) (Sakkoulas ed) 54, 107, 115, 462-469.
- 452 Papadimitrakis, G (2019), 'Big data και αλγοριθμικές μελέτες επικινδυνότητας – Νέες προκλήσεις στον χώρο της ποινολογίας' (Big data and algorithmic risk research – New challenges in the field of penal sentences) 10 Ποινική Δικαιοσύνη (*Criminal Justice*) 1045-1054.
- 453 Skondra, M (2020), 'Αναγνώριση προσώπου (Facial recognition) και προσωπικά δεδομένα' (Facial recognition and personal data) (*Ddikastes*, 20 January 2020) available at: www.ddikastes.gr/node/5271.
- 454 Leftheriotou, E (2019), 'Οι Προκλήσεις της Ρομποτικής και της Τεχνητής Νοημοσύνης για το Νομοθέτη και το Δικαστή' (The challenges of robotics and artificial intelligence for the legislator and the judge) in Delouka-Igglessi, K, Lignomenou, A, Sinanioti-Marouda, A, Δίκαιο και Τεχνολογία (*Law and Technology*) (Sakkoulas ed).
- 455 Papadimitrakis, G (2019), 'Big data και αλγοριθμικές μελέτες επικινδυνότητας – Νέες προκλήσεις στον χώρο της ποινολογίας' (Big data and algorithmic risk research – New challenges in the field of penal sentences) 10 Ποινική Δικαιοσύνη (*Criminal Justice*) 1045-1054.
- 456 See www.vb.is/frettir/raeda-stodu-kvenna-i-upplysingataekni/154387/; www.si.is/frettasafn/fundur-um-innbyggda-fordoma-i-gervigreind/; www.mbl.is/frettir/innlent/2019/06/19/ny_stjorn_samtaka_kvenna_i_upplysingataekni/; www.youtube.com/watch?v=afwfuXhqlGQ.
- 457 Leavy, S (2018), 'Gender Bias in Artificial Intelligence: the need for diversity and gender theory in machine learning', 2018 IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE); see also e.g. Foale, N (2020), 'Back to the Future: how well equipped is Irish employment equality law to adapt to artificial intelligence?' 23(1) *Trinity College Law Review* 170.
- 458 See <http://www.williamfry.com/newsandinsights/news-article/2018/04/25/artificial-intelligence-in-the-workplace-an-employment-law-perspective> and [https://www.williamfry.com/newsandinsights/news-article/2018/05/31/artificial-intelligence-in-the-workplace-\(part-2\)-ai-assisted-recruitment-and-employment-equality-law](https://www.williamfry.com/newsandinsights/news-article/2018/05/31/artificial-intelligence-in-the-workplace-(part-2)-ai-assisted-recruitment-and-employment-equality-law).
- 459 See www.unipi.it/index.php/risultati-e-prodotti/itemlist/category/1548-salvatore-ruggieri.
- 460 In **Lithuania**, an interdisciplinary seminar entitled 'Judge, decision-making and artificial intelligence' on the potential usage of algorithms in the judicial system was organised by Vilnius University Tech Hub, Vilnius University Faculty of Law in December 2019; see www.tf.vu.lt/ivykiai/elektroninis-asmuo-ateities-butinybe-ar-tik-fikcija/. PhD projects include Aurelija Šerniūtė's research at Vilnius University Faculty of Law on 'Non-discriminatory artificial intelligence – mission impossible?'; presented at the Geneva Digital Law Research Colloquium (19 June 2020; publication in progress); see www.unige.ch/droit/pi/files/5515/9601/1869/Aurelija_Serniute.pdf.

and European law.⁴⁶¹ The creation of a joint AI laboratory at the University of **Luxembourg** and the Luxembourg Institute of Science and Technology co-sponsored by the Luxembourgish Government and the private tech company NVIDIA in 2019 also signals the emergence of scientific discussions on AI and the law, including on issues of algorithmic discrimination.⁴⁶² In **Norway**, although much of the legal scholarship has focused on data protection issues, some attention has been granted to issues of algorithmic bias in particular in relation to the use of AI in health and biotechnology.⁴⁶³ The **Portuguese** expert has reported that legal scholarship on the impact of algorithms on gender equality and non-discrimination remains very limited, although some mapping of the issue has been conducted in relation to labour law.⁴⁶⁴ In **Spain**, the scientific community is showing increasing interest in the ethical problems arising from the use of AI, including algorithmic discrimination.⁴⁶⁵ For example, the first 'Workshop on ethics in artificial intelligence: searching for an ethical algorithm', was held in Madrid in April 2019 to discuss the social impact of AI and potential strategies to foster the positive impact of AI.⁴⁶⁶ Nevertheless, most of the available scholarship focuses on violations of rights such as privacy or general principles of

-
- 461 In **Sweden**, a thesis from Umeå University entitled 'Automated decision-making vs indirect discrimination: Solution or aggravation?' was presented in 2019; see Lundberg, E (2019), 'Automated decision-making vs indirect discrimination. Solution or aggravation?' (Master of Science thesis, Umeå University) available at: www.diva-portal.org/smash/record.jsf?pid=diva2%3A1331907&dsid=6537. The author argues that a greater legal focus on the alleged harm to the applicant in indirect discrimination cases rather than on her/his traits or on finding an adequate comparator could be a way to better capture discrimination caused by automated decision-making systems in courts. The thesis analyses various legal contexts: the US, Canada and Sweden as well as the European Court of Human Rights and the Court of Justice of the European Union. Another Master thesis from Lund University entitled 'Machine Bias: Artificial Intelligence and Discrimination' and published in 2019 notes that the most vulnerable social groups are more likely to be victims of algorithmic discrimination and explores regulatory solutions. This thesis focuses on the US, the European Union and China, see Yavuz, C (2019), 'Machine Bias: Artificial Intelligence and Discrimination' (LLM thesis, Lund University) 4, available at: <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=8987035&fileId=8987040>.
- 462 A University of Luxembourg researcher for instance explained: 'At a later stage, I am interested in investigating bias data and fairness in AI to address disparity in model performance particularly between genders', see: <https://wide.lu/interview-with-amal-tawakuli/>.
- 463 Research projects include: Befring, AK (2019), *Persontilpasset medisin. Rettslige perspektiver* (Personalised medicine. Legal perspectives) (Gyldendal) available at: www.gyldendal.no/Faglitteratur/Jus/Juridiske-fag/Persontilpasset-medisin; BigMed, an interdisciplinary programme aiming at developing a methodology for handling big data in the medical field: bigmed.no; an interdisciplinary project on machine learning in the medical sector at Tromsø university, see https://uit.no/nyheter/artikkel?p_document_id=662276. In addition, research on law and digitalisation has touched on various other issues such as human rights and AI in judicial and public decision making, see Langford, M (2020), 'Taming the Digital Leviathan: Automated Decision-Making and International Human Rights' 114 *American Journal of International Law Unbound* 141 available at: www.cambridge.org/core/journals/american-journal-of-international-law/article/taming-the-digital-leviathan-automated-decisionmaking-and-international-human-rights/5AFE96F03A1B75B63729D60F0F609609. In relation to predictive policing and machine learning, see the research conducted by postdoc Mareile Kaufman at the Institute for Criminology and Sociology of the Law at University of Oslo (www.jus.uio.no/ikrs/personer/vit/mareilek/index.html); on eGovernment, data protection, legislative drafting and legal technology, see the work of Professor Dag Wiese Scharthum at the Institute for Private Law at the University of Oslo (www.jus.uio.no/ifp/personer/vit/dags/); on IT and universal design, see associate professor Jo Herstad's research conducted at the Department of Informatics at the University of Oslo (www.mn.uio.no/ifi/personer/vit/johe/index.html).
- 464 Moreira, TC (2019), 'Igualdade de Género no Trabalho 4.0' (Gender equality in 4.0 Work) in Palma Ramalho, MR and Moreira, TC, *A Igualdade nas Relações de Trabalho* (Equality in Employment Relations) (Estudos Apodit 5, AAFDL Editions, Lisbon) 45-68; Moreira, TC (forthcoming), 'Algorithms discrimination and Labour Law' (forthcoming) in *Anuário de Direitos Humanos* No. 2.
- 465 The Eticas Foundation has stressed the need for applying multidisciplinary approaches, considering algorithmic processes 'as technological processes integrated in particular contexts and specific practices'; see Eticas Foundation (2018), 'Algorithmic discrimination. One of the main challenges for social progress in the 21st century' (23 March 2018), available at: <https://eticasfoundation.org/algorithmic-discrimination-one-of-the-main-challenges-for-social-progress-in-the-21st-century/>.
- 466 The recording and conclusions of the workshop can be consulted here: González-Espejo García, MJ (2019), 'Video y conclusiones de las "I Jornadas sobre Ética e Inteligencia Artificial: en busca de un algoritmo ético"' (Video and conclusions of the 1st Workshop on Ethics and Artificial Intelligence: in search of an ethical algorithm) University CEU San Pablo (Madrid), 2 April 2019 Institute of Legal Innovation, available at: <https://institutoeinnovacionlegal.com/jornadas-etica-e-ia/>.

criminal law. Academic work to date has not yet focused on the specific implications of the problem of algorithmic discrimination for the Spanish gender equality and non-discrimination legal framework.⁴⁶⁷

Beyond these instances of emerging exploratory research on the discriminatory impact of AI and the challenges to national gender equality and non-discrimination law, several national research institutions have launched scientific projects on algorithmic discrimination in a range of domains and often in an interdisciplinary fashion. For instance, in **Czechia**, where the topic of algorithmic discrimination is emerging in national research,⁴⁶⁸ several scientific institutions (the Academy of Science, the Czech Technical University, Charles University) and the City of Prague, in collaboration with private firms, have launched a project called prg.ai, which aims to research topics such as the reasonable regulation of AI and to 'support discussion on the ethical, legal and socio-economical aspects of AI', notably through funding a research position.⁴⁶⁹ In **Denmark**, research on algorithmic discrimination has for example focused on the consequences of bias in medical AI,⁴⁷⁰ algorithmic bias in public administration⁴⁷¹ and discrimination in the labour market where AI is used in recruitment processes⁴⁷² and targeted advertising.⁴⁷³ In **Germany**, existing legal scholarship on algorithmic discrimination has highlighted the lack of fit between current non-discrimination law and the discriminatory harms produced by algorithmic discrimination.⁴⁷⁴ The German Women Lawyers' Association, responding to the Commission's white paper on artificial intelligence, has highlighted that non-discrimination law only captures discriminatory decisions based on algorithmic classifications as opposed to these classifications themselves, in spite of the fact that they can be problematic, for example if they force subjects into binary gender identity categories.⁴⁷⁵ Another friction concerns the applicability of non-discrimination law to the provision of goods and services that are

- 467 E.g. Cotino Hueso, L (2017), 'Big Data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales' (Big data and artificial intelligence. an approach from a legal point of view about fundamental rights) 24 *Dilemata* 131; Astobiza, AM (2017), 'Ética algorítmica: Implicaciones éticas de una sociedad cada vez más gobernada por algoritmos' (Algorithmic Ethics: Ethical Implications of a Society Increasingly Governed by Algorithms) 24 *Dilemata* 185; Miró Llinares, F (2018), 'Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots' (Artificial Intelligence and criminal justice: beyond the harmful results caused by machines) 20 *Revista de Derecho Penal y Criminología* 87; Delépine, J (2019), 'Algoritmos y discriminación sexista' (Algorithms and sexist discrimination) 69 *Alternativas económicas* 23; Benítez Eyzaguirre, L (2019), 'Ética y transparencia para la detección de sesgos algorítmicos de género' (Ethics and transparency for detection of gender bias in algorithms) 25(3) *Estudios sobre el mensaje periodístico* 1307; Lazcoz Moratino, G (forthcoming, on file with the author), 'Modelos algorítmicos, sesgos y discriminación' (Algorithmic models, bias and discrimination).
- 468 See e.g. the monograph by Lobotka, A (2019), *Umělá inteligence z pohledu antidiskriminačního práva a GDPR* (Artificial intelligence from the point of view of the antidiscrimination law and GDPR) (Wolters Kluwer); Hanych, M (2019), 'Umělá inteligence a lidská práva – nevyhnutelný střet?' (Artificial intelligence and human rights – the inevitable clash) 10 *Bulletin centra pro lidská práva a demokratizaci* (Bulletin of the Center for Human Rights and Democratisation) 9; Hoder, L (2020), 'Jan Lhotský: Umělá inteligence a lidská práva – obor budoucnosti?' (Jan Lhotský: Artificial intelligence and human rights – a field of the future) *Jiné Právo*, (2 April 2020) available at: <https://jinepravoblogspot.com/2020/04/jan-lhotsky-umela-inteligence-lidska.html>.
- 469 Available at: <https://prg.ai/en/>.
- 470 One such multidisciplinary project is the AI@CARE project, gathering computer scientists and lawyers, which started on 1 April 2020 and will last for three years, and is funded by the University of Copenhagen's DATA+ pool, see University of Copenhagen (2020) 'New project will help to prevent bias and discrimination in medical AI' (May 2020) available at: <https://di.ku.dk/english/news/2020/new-project-will-help-to-prevent-bias-and-discrimination-in-medical-ai/>.
- 471 See for instance the Public Administration and Computational Transparency in Algorithms – PACTA project, which brings together lawyers and computer scientists and is also concerned with issues of discrimination, see 'Public Administration and Computational Transparency in Algorithms – PACTA' (University of Copenhagen) available at: <https://jura.ku.dk/icourts/research/pacta/>.
- 472 See e.g. a research grant awarded by the Aarhus University Research Foundation in 2019 expected to be carried out in 2020-2022: Bagger, T (2019), 'AUFF starting grant of 2,200,000 kr. for Vincenzo Pietrogiovanni from Department of Law' *Juridisk Institut*, (19 December 2019) available at: <https://law.medarbejdere.au.dk/aktuelt/nyheder/nyhed/artikel/auff-starting-grant-of-2200000-kr-to-vincenzo-pietrogiovanni-from-department-of-law/>.
- 473 See e.g. Schjøler, CH (2018), 'Can the use of online targeted advertising in job adverts constitute age discrimination?' (Colloquium Fundamental Rights in the Workplace, 27-28 November 2018, Aarhus University) available at: <https://portal.findresearcher.sdu.dk/en/publications/can-the-use-of-online-targeted-advertising-in-job-adverts-constit>.
- 474 See Fröhlich, W and Spiecker genannt Döhmman, I (2018), 'Können Algorithmen diskriminieren?' (Can algorithms discriminate?) *Verfassungsblog*, (26 December 2018) available at: verfassungsblog.de/koennen-algorithmen-diskriminieren/.
- 475 German Women Lawyers' Association (2020), 'Stellungnahme zu dem Weißbuch der EU-Kommission "Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen"' (Statement on the EU Commission white paper on artificial intelligence – a European concept for excellence and trust) (Berlin, 9 June 2020) available at: www.djb.de/verein/Kom-u-AS/ASDigi/st20-20/.

distributed in bulk but have been personalised by algorithms.⁴⁷⁶ German scholars have further addressed problems of liability for algorithmic discrimination and enforcement of non-discrimination law.⁴⁷⁷ In the **Netherlands**, the legal challenges raised by algorithmic discrimination have been identified early on in a report dealing with the broader question of algorithms and fundamental rights.⁴⁷⁸ In addition, a recent major study has analysed the discrimination risks of the use of algorithmic decision making in a number of sectors, such as the judiciary, and related to specific topics, such as content moderation by social media and automated cars.⁴⁷⁹ Academics have further specifically examined algorithmic discrimination in the context of platform work for the Netherlands, the collaborative economy and the labour market in general, highlighting problems linked to the enforcement of non-discrimination law, in particular in relation to the burden of proof, in the context of a lack of transparency and explainability of algorithmic outputs.⁴⁸⁰ In the **United Kingdom**, the national expert has reported concerns expressed by academics over both the ability of a claimant to identify algorithmic discrimination and the ability of a defendant to provide legal justification for algorithmic decisions in the algorithmic 'black box'.⁴⁸¹ Scholars also point out that the academic debate in the UK has not yet extensively explored the problem of the unsuitability of existing legal tests for algorithmic discrimination.⁴⁸² Recently, British experts have examined the role that European equality bodies can play in monitoring and addressing algorithmic discrimination.⁴⁸³

3.3.2.2 Interactions between data protection law and gender equality and non-discrimination law in national scholarship

For most countries (**Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Hungary, Iceland, Ireland, Latvia, Liechtenstein, Lithuania, Malta, Netherlands, Romania, Slovakia, Slovenia, Sweden**), national experts have reported limited or no legal scholarship analysing the potential interaction between national non-discrimination and data protection regulations in addressing issues of algorithmic discrimination.

-
- 476 Busch, C (2018), *Algorithmic Accountability* (Expert report in the framework of the 'ABIDA – Assessing Big Data' project funded by the German Federal Ministry for Education and Research) available at: www.abida.de/sites/default/files/ABIDA%20Gutachten%20Algorithmic%20Accountability.pdf, 46ff. In fact, EU non-discrimination law is only applicable to goods and services that are 'available to the public' and the algorithmic personalisation of offers questions this notion of public availability.
- 477 See De Barros Fritz, R and Von Lewinsky, K (2018), 'Arbeitgeberhaftung nach dem AGG infolge des Einsatzes von Algorithmen bei Personalentscheidungen' (Employer liability due to the use of algorithms in personnel decisions) *Neue Zeitschrift für Arbeitsrecht* 620; Scheer, J (2019), *Algorithmen und ihr Diskriminierungsrisiko. Eine erste Bestandsaufnahme* (Algorithms and their Risk of Discrimination. A first stocktaking), State Anti-Discrimination Body of Berlin, available at: www.berlin.de/sen/lads/ueber-uns/materialien/, 14; see also: www.berlin.de/sen/lads/assets/ueber-uns/materialien/algorithmenDiskriminierungsrisiko_bf.pdf.
- 478 Vetz, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten* (Algorithms and fundamental rights) (Den Haag, Boom Juridisch).
- 479 Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek* (Legal aspects of algorithms that make decisions. An exploratory study) (WODC/Utrecht University).
- 480 Gerards, J (2019), 'Grondrechten in de platformeconomie' (Fundamental rights in the platform economy) in Gerards, J and Van Schaik, A, *Preadvies Vereniging voor Burgerlijk Recht* (Pre-Advice for the Netherlands Association of Private Law) (Uitgeverij Paris, Amsterdam); Jovović, M (2018), 'Ter Visie – Algoritmische discriminatie' (For Vision – Algorithmic Discrimination) 4 *Arbeid en Onderneming* (Employment and Enterprise) 141; Kullmann, M (2018), 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law', p. 1; Kullmann, M (2019), 'Platformwerk, besluitvorming door algoritmen en bewijs van algoritmische discriminatie' (Platform work, algorithmic decision-making and proof of algorithmic discrimination) 8 *Ondernemingsrecht* (Business Law); Burri, S and Heeger-Hertter, S (2018), 'Discriminatie in de platformeconomie juridisch bestrijden: geen eenvoudige zaak' (Combating discrimination in the platform economy with legal means: No simple matter) *Ars Aequi* 1000.
- 481 See Allen, R and Masters, D (2020), 'Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making' 20(4) *ERA Forum* 585.
- 482 Allen, R (2020), 'Artificial intelligence, machine learning, algorithms and discrimination law: the new frontier' (Discrimination Law Conference, Michael Rubenstein Conferences, London, 31 January 2020) available at: https://482pe539799u3ynseg2hl1r3-wpengine.netdna-ssl.com/wp-content/uploads/2020/02/Discrimination-Law-in-2020_FINAL_-1.pdf.
- 483 Allen, R and Masters, D (2020), *Regulating for an equal AI: A New Role for Equality Bodies. Meeting the new challenges to equality and non-discrimination from increased digitisation and the use of Artificial Intelligence* (Equinet, Brussels) available at: https://equineteurope.org/wp-content/uploads/2020/06/ai_report_digital.pdf.

By contrast, the potential strengths and weaknesses of combining these regulatory tools have been highlighted in countries such as **Germany, Greece, Norway, Poland** and the **United Kingdom**. For example, **German** scholars have attracted attention to the insufficiencies of the GDPR and national data protection law when it comes to questions of equality and non-discrimination due to the limited scope of application of transparency and accountability obligations, and have argued for complementary regulatory tools beyond the data protection regulations.⁴⁸⁴ In **Greece**, generally speaking, the current debate on algorithmic discrimination has been conducted through the lens of data protection rather than gender equality and non-discrimination law, but the interaction between the two regulatory bodies has also been the subject of a few research projects.⁴⁸⁵ It is argued in such studies that the assessment of compliance with ethical and social values is more complicated than the 'traditional' data protection assessment but also that transparency, intelligibility, and explainability are crucial principles for the purposes of gender equality and non-discrimination.⁴⁸⁶ Combining insights from data protection and non-discrimination law, other studies emphasise the need for different levels of human control,⁴⁸⁷ the right to informative self-determination of algorithmic subjects⁴⁸⁸ and the discriminatory risks ensuing from data processing and classification in the absence of human intervention.⁴⁸⁹ In **Norway**, the principles regulating the fair processing of data contained in Article 5(1) GDPR have been described by scholars as the main legal instrument for equality issues but have also been criticised for not offering sufficiently practical guidelines to practitioners.⁴⁹⁰ In addition, provisions on access to information have been criticised for offering an either/or approach to information (classified as either sensitive or not sensitive) instead of grading the sensitivity of the information and the subsequent access to it in a more granular fashion.⁴⁹¹ In the **United Kingdom**, scientific literature has largely focused on the right to transparency and existing correlated gaps in relation to the right to understand automated decision making. Researchers have also examined the potential and feasibility of a 'right to explanation' that could enhance the right to transparency.⁴⁹² Some attention has been granted to the interaction of the regulatory fields of data protection law and non-

-
- 484 Dreyer, S and Schulz, W (2018), *Was bringt die Datenschutzgrundverordnung für automatisierte Entscheidungssysteme? Potenziale und Grenzen der Absicherung individueller, gruppenbezogener und gesellschaftlicher Interessen* (What are the benefits of the basic data protection regulation for automated decision systems? Potentials and limits of safeguarding individual, group-related and social interests) (Bertelsmann Stiftung ed) available at: https://algorithmenethik.de/wp-content/uploads/sites/10/2018/04/BSt_DSGVOundADM_dt_ohneCover.pdf.
- 485 Professor Mitrou's 'Data Protection, Artificial Intelligence and Cognitive Services' deals extensively with algorithmic discrimination and data protection law: Mitrou, L (2018), 'Data Protection, Artificial Intelligence and Cognitive Services. Is the General Data Protection Regulation (GDPR) "Artificial Intelligence-Proof?"' (a study commissioned by Microsoft but expressing the views and opinions of the author), available at: <https://ssrn.com/abstract=3386914>.
- 486 Mitrou, L (2018), 'Data Protection, Artificial Intelligence and Cognitive Services. Is the General Data Protection Regulation (GDPR) "Artificial Intelligence-Proof?"'.
- 487 Vorras, A and Mitrou, L (2018), 'Τεχνητή νοημοσύνη και προσωπικά δεδομένα – Μια θεώρηση υπό το πρίσμα του Ευρωπαϊκού Γενικού Κανονισμού Προστασίας Δεδομένων (ΕΕ) 2016/679' (Artificial Intelligence and personal data – An overview under the prism of the European General Data Protection Regulation (EU) 2016/679) 4 *Δίκαιο Μέσων Ενημέρωσης & Επικοινωνίας* (ΔιΜΕΕ) 460.
- 488 Douka, V (2018), 'Προσωπικά δεδομένα και εργασιακές σχέσεις: προστασία σύμφωνα με τον Γενικό Κανονισμό Προστασίας Δεδομένων' (Personal data and employment relationships: Protection according to the GDPR) 6 *Επιθεώρησης Εργατικού Δικαίου* (ΕΕργΔ) (*Labour Law Review*) 641; Ladas, D (2018), Η προσωπικότητα του εργαζομένου (The personality of the employee) (Athens, Nomiki Vivliothiki) 89ff.
- 489 Panagopoulou-Koutnatzi, F (2017), 'Ο Γενικός Κανονισμός για την Προστασία Δεδομένων 679/2016/ΕΕ' (The General Data Protection Regulation 679/2016/EU) (Sakoulas ed) 112; see also Panagopoulou-Koutnatzi, F (2017), 'Τα νέα δικαιώματα για τους πολίτες βάσει του Γενικού Κανονισμού Προστασίας Δεδομένων: μια πρώτη αποτίμηση και συνταγματική αξιολόγηση' (The new rights of citizens under the GDPR: a first assessment and constitutional evaluation) *Εφημερίδα Διοικητικού Δικαίου* (ΕΔΔ) (*Administrative Law Journal*) 81; Pertsinidou, A (2019), 'Η κατάρτιση εξατομικευμένων προφίλ ως εργαλείο αξιολόγησης στο ψηφιακό περιβάλλον εργασίας. Εξόρυξη γνώσης από βάσεις δεδομένων: Κίνδυνοι και προκλήσεις' (Profiling as an evaluation tool in the digital labour environment. Data mining: risks and challenges) 8 *Επιθεώρησης Εργατικού Δικαίου* (ΕΕργΔ) (*Labour Law Review*) 931; Sotiropoulos, V (2019), Υπεύθυνος Προστασίας προσωπικών Δεδομένων (Data Protection Officer) (Sakkoulas ed, 2nd edn) 185; Hatzopoulos, V (2020), 'Η διακυβέρνηση του κυβερνοχώρου' (The governance of the cyberspace) in Stephanou, K (ed) (2020), Το σύστημα διακυβέρνησης της Ευρωπαϊκής Ένωσης (The system of governance of the European Union) (Athens, Nomiki Vivliothiki) 402.
- 490 Interview conducted by Norwegian expert Marte Bauge with Prof Schartum (4 June 2020).
- 491 Befring, AK (2019), *Persontilpasset medisin. Rettslige perspektiver* (Personalised medicine. Legal perspectives) (Gyldendal), available at: www.gyldendal.no/Faglitteratur/Jus/Juridiske-fag/Persontilpasset-medisin.
- 492 Authors have doubted the legal existence and feasibility of such a right, see e.g. 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*, available at: <https://ssrn.com/abstract=2903469> or <http://dx.doi.org/10.2139/ssrn.2903469>.

discrimination law in **France**,⁴⁹³ **Italy**,⁴⁹⁴ **Luxembourg**,⁴⁹⁵ **Poland**⁴⁹⁶ and **Portugal**.⁴⁹⁷ French scholarship has also criticised the limits of the GDPR in tackling issues of algorithmic discrimination: although data protection and anti-discrimination law can pursue common purposes, more specific obligations are needed to prevent systemic discrimination and determine how to make stakeholders accountable.⁴⁹⁸

As mentioned above and as further discussed in section 3.4.1.1, the GDPR is thus considered a possible avenue for combating algorithmic discrimination, but it is also considerably criticised for its limited scope of application, for instance in relation to partially automatised decisions. As Chapter 4 explains, other legal tools beyond data protection law have also been highlighted by scholars as potentially complementary to anti-discrimination law, such as competition law⁴⁹⁹ or intellectual property law.

3.3.2.3 Issues of liability and responsibility for algorithmic discrimination

While national experts report rich discussions on general questions of responsibility and liability in the context of algorithmic decision making and on the legal personhood of AI, legal scholarship on the specific question of liability for algorithmic discrimination appears limited in a majority of countries (**Belgium, Bulgaria, Croatia, Cyprus, Denmark, Estonia, Greece, Hungary**,⁵⁰⁰ **Iceland, Ireland, Latvia, Liechtenstein, Lithuania, Luxembourg, Portugal, Malta, Norway, Poland, Romania, Slovakia, Slovenia, Sweden**). This seems a logical consequence of the scarcity of scientific discussions on algorithmic discrimination as such.

The question has nonetheless attracted scientific attention in some countries. In **Austria**, for example, the research undertaken in relation to the AMS algorithm described above has highlighted concerns over the lack of opportunity for users of the algorithm to seek remedies in relation to its decisions, and

493 Le Clainche, J and Le Métayer, D (2012), 'Données personnelles, vie privée et non-discrimination : des protections complémentaires, une convergence nécessaire' (Personal data, privacy and non-discrimination: complementary protections, a necessary convergence), *Revue Lamy Droit immatériel* 90; Mercat-Bruns, M (ed) (2020), *Nouveaux modes de detection et de prevention de la discrimination et accès au droit : action de groupe et discrimination systémique, algorithmes et préjugés, réseaux sociaux et harcèlement* (New ways of detecting and preventing discrimination and access to the law. Class action and systemic discrimination. Algorithms and prejudices. Social networks and harassment) (Société de législation comparée, Collection Trans Europe Experts) available at: www.lgdj.fr/nouveaux-modes-de-detection-et-de-prevention-de-la-discrimination-et-acces-au-droit-9782365170925.html.

494 See <https://www.forumdisuguaglianzediversita.org/wp-content/uploads/2019/09/Resta.x89907.pdf>.

495 During the annual Conference 2019 of CPDP in Brussels (30 January to 1 February 2019), the Interdisciplinary Centre for Security, Reliability and Trust (SnT) of the University of Luxembourg organised a panel entitled 'AI Governance. Role of legislators, tech companies and standard bodies'; Alain Herrman (National Commission for Data Protection) gave a presentation on 'Artificial Intelligence Governance. Point of view of DPA'. See the conference website: <https://cpdpconferences.net/CPDP2019.pdf> and see also <https://cnpd.public.lu/dam-assets/fr/actualites/international/2019/cdpdp2019-presentation-CNPD.pdf>.

496 Niklas, J, Sztandar-Sztanderska, K and Szymielewicz, K (2015), *Profiling the unemployed in Poland: social and political implications of algorithmic decision making* (Eliza Jakubiak tr, report by Fundacja Panoptykon, Warsaw) available at: https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf.

497 As in the study of Moreira, T (2019), 'Igualdade de Género no Trabalho 4.0' (Gender equality in 4.0 Work) in Palma Ramalho MT and Coelho Moreira, T (eds), *A Igualdade nas Relações de Trabalho* (Equality in Employment Relations) (Estudos APODIT 5, Lisbon (AAF DL Editions)) 66-7.

498 See the articles of Porta, JT and others in Mercat-Bruns, M (ed) (2020), *Nouveaux modes de detection et de prevention de la discrimination et accès au droit : action de groupe et discrimination systémique, algorithmes et préjugés, réseaux sociaux et harcèlement* (New ways of detecting and preventing discrimination and access to the law. Class action and systemic discrimination. Algorithms and prejudices. Social networks and harassment) (Société de législation comparée, Collection Trans Europe Experts) available at: www.lgdj.fr/nouveaux-modes-de-detection-et-de-prevention-de-la-discrimination-et-acces-au-droit-9782365170925.html.

499 Chellet, P (2020), 'Prix personnalisé sur internet : la rencontre de la discrimination, de la concurrence et des données personnelles', in Mercat-Bruns, M (ed), *Nouveaux modes de detection et de prevention de la discrimination et accès au droit : action de groupe et discrimination systémique, algorithmes et préjugés, réseaux sociaux et harcèlement*, 103.

500 One of the only contributions on this issue is in relation to possible problems regarding the enforcement of norms, see Zódi, Zs (2017): 'Az információs társadalom legújabb kihívásai a jog számára – Horizontális platformok' (The newest challenges for the law posed by the information society – Horizontal platforms), *Gazdaság és Jog* (No 9) 28. The author refers to previous experiences when the attempts of the Hungarian Government to make Google and Facebook (ie, key actors in the field) pay advertisement tax in Hungary failed because these companies were not considered to come under national jurisdiction.

posed the question of a possible right to appeal algorithmic decisions.⁵⁰¹ The lack of sufficient legal remedies in relation to the decisions of the AMS algorithm was one of the reasons, along with the lack of legal basis for profiling certain groups of population, that led the **Austrian** data protection agency (*Datenschutzbehörde* or *DSB*), relying *inter alia* on this scientific research, to issuing a prohibitive decree in 2020 regarding the use of the algorithm.⁵⁰² The **Czech** working group on human rights and new technologies under the Office of the Government for Human Rights has also started to explore issues of responsibility and liability for algorithmic discrimination.⁵⁰³ Turning to the employment context, **Dutch** scholars have argued that employers may be liable for discriminatory outcomes of AI systems in the employment context if an automated decision-making system is not transparent and they are unable to explain the reasons for differential outcomes.⁵⁰⁴ Dutch literature has differentiated between various kinds of liability for algorithmic discrimination.⁵⁰⁵ Designers of discriminatory algorithms could be held responsible following principles of tort liability (product or owner liability) for creating a malfunctioning algorithm and breaching their duty of care, for example if they failed to build limits into the system or create opportunities to monitor or correct algorithmic systems. By contrast, risk liability would capture an employer's liability for discrimination resulting from the use of algorithms by its employees, both against third parties and against its own employees. Civil liability is a third kind of liability that could help capture situations of algorithmic discrimination. Concluding that Dutch law is sufficiently equipped to capture liability for algorithmic discrimination, the author of an extensive study on algorithmic liability balances arguments for a strict liability (an economic rationale solution following which the party who is most able to prevent the damage has to pay for it when it arises)⁵⁰⁶ against arguments for a more lenient approach to liability for algorithms (based on arguments that a strict liability regime would hinder innovation).⁵⁰⁷ In turn, **French** scholars have highlighted the lack of clarity on responsibilities for algorithmic discrimination, especially in case of machine-learning and deep-learning algorithms, and have argued that the lack of a clear legal framework increases the need for information, monitoring and testing beyond judicial scrutiny.⁵⁰⁸ Others have argued that intellectual property rights can undermine transparency rules and complicate the enforcement of responsibility and liability rules.⁵⁰⁹ In addition, some French scholars have doubted the ability of existing remedies such as class action to capture liability for algorithmic discrimination.⁵¹⁰ Based on an analogy with the principle of 'privacy by design', a

501 See Allhutter, D and others (2020), 'Algorithmic profiling of job seekers in Austria: how austerity politics are made effective', *Frontiers in Big Data*, available at: <https://www.frontiersin.org/articles/10.3389/fdata.2020.00005/full>.

502 See Staudacher, A (2020), 'Datenschutzbehörde dreht den AMS-Algorithmus ab' (Data protection authority turns off the AMS algorithm), *Kurier* (20 August 2020) available at: <https://kurier.at/wirtschaft/datenschutzbehoerde-dreht-den-ams-algorithmus-ab/401006615>; Szigetvari, A (2020), 'Datenschutzbehörde kippt umstrittenen AMS-Algorithmus' (Data protection authority overturns controversial AMS algorithm), *Der Standard* (20 August 2020) available at: <https://www.derstandard.at/story/2000119486931/datenschutzbehoerde-kippt-umstrittenen-ams-algorithmus>.

503 In 2019, this working group published a recommendation for the Digital Czechia programme, which also touched upon discrimination, even if in a general way. See Working Group for Human Rights and Modern Technologies of the Council of the Government of the Czech Republic for Human Rights, available at: www.vlada.cz/cz/ppov/rfp/vybory/pro_lidska-prava-a-modern-technologie/stranka-se-pripravuje-176819/.

504 Jovović M. (2018), 'Ter Visie – Algoritmische discriminatie' (For Vision – Algorithmic Discrimination), *Arbeid en Onderneming*, 141.

505 Tjong Tjin Tai, E (2017), 'Aansprakelijkheid voor robots en algoritmes' (Liability for robots and algorithms), 14(3) *Nederlands Tijdschrift voor Handelsrecht*, 123.

506 This is by analogy with a 'Bundesgerichtshof' judgment in which the outcome of the Google auto-complete function was considered to be Google's own content, even though it had been created automatically; see BGH, 14 May 2013, VI ZR 269/12 (*Google Auto-complete*). See about this judgment: Husovec, M (2013), 'Liability for Algorithm Design & Big Data', available at: www.husovec.eu/2013/06/liability-for-algorithm-design-big-data.html.

507 Tjong Tjin Tai, E (2017), 'Aansprakelijkheid voor robots en algoritmes' (Liability for robots and algorithms), 14(3) *Nederlands Tijdschrift voor Handelsrecht*, section 9. See also Breemen, VE and Wouters AHH (2009), 'Casestudy Zelfrijdende auto's' (Case study automated cars), in Kulk, S, van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek* (Legal aspects of algorithms that make decisions. An exploratory study) (The Hague, WODC).

508 Cluzel L (2020), 'Décision publique algorithmique et discrimination' in Mercat-Bruns, M (ed), *Nouveaux modes de detection et de prevention de la discrimination et accès au droit: action de groupe et discrimination systémique, algorithmes et préjugés, réseaux sociaux et harcèlement*, 107.

509 Le Clainche J and Le Métayer D (2013), 'Données personnelles, vie privée et non-discrimination: des protections complémentaires, une convergence nécessaire' (Personal data, privacy and non-discrimination: complementary protections, a necessary convergence) 90 *Revue Lamy Droit immatériel*.

510 Porta, J (2020), 'Algorithme et risques discriminatoires' in Mercat-Bruns, M (ed), *Nouveaux modes de detection et de prevention de la discrimination et accès au droit: action de groupe et discrimination systémique, algorithmes et préjugés, réseaux sociaux et harcèlement*, 68.

French scholar has proposed the concept of ‘responsibility by design’, extending liability for algorithmic discrimination to those who conceive algorithms.⁵¹¹ In **Germany**, scholars have underlined the structural inequalities consumers face as they have no real ability to detect, let alone prove, infringements of non-discrimination law in algorithm-based systems.⁵¹² To ease the enforcement of non-discrimination law, a German scholar suggests that ‘[i]n view of [the existing] structural [information] asymmetry, courts should accommodate the user of personality-sensitive software applications in the liability process with a differentiated [or graded] system of burden of proof distribution in accordance with the principle of procedural equality’.⁵¹³ He proposes that it should suffice that the subject of algorithmic decisions present facts from which it can be presumed that algorithmic discrimination has arisen. It should then be for the defendant, i.e. the provider of the algorithmic application, to offer counter-evidence by means of ‘logged program sequences, proof of sufficient supervision of the technical processes used or by otherwise demonstrating the absence of causality’ between the application and the discrimination at stake.⁵¹⁴ The author also argues that a fair distribution of liability would mean that the provider of the algorithmic application should be held liable if ‘there are indications that the error was detectable’, which in this case would mean that the discrimination could have been avoided.⁵¹⁵ In addition, he argues that a strict liability, paired with an insurance obligation to ensure solvability in case of damages, should be established in sensitive areas such as healthcare, based on the rationale that those who profit from software applications should also be liable for them and that liability should be adapted to the sensitivity of the area of application.⁵¹⁶ In the employment context, German scholars have suggested that while employers should be held liable for algorithmic discrimination, technical ‘anti-discrimination by design’ solutions would be more timely and effective than legal solutions, which would require time to be put in place.⁵¹⁷ In **Italy**, this question has been mainly examined in the context of the use of algorithms in public administration. Academics have argued that responsibility rules should be adapted and have proposed the idea of a double responsibility following which (1) the official who acted on behalf of the public administration should be held responsible, and (2) responsibility should also lie with the designers of the algorithm.⁵¹⁸

3.4 Legal responses to algorithmic discrimination in the European countries

3.4.1 Legislative instruments

3.4.1.1 Relevant legislative instruments in the European countries

Limited specific legislation concerning algorithmic discrimination

Section 3.1 has shown that algorithms are frequently used in many European countries, both in the public and in the private sector, and sections 3.2 and 3.3 have confirmed that this is seen to present specific challenges and risks of discrimination: algorithms may be fed biased data and may have discriminatory

511 Cluzel L (2020), ‘Décision publique algorithmique et discrimination’ in Mercat-Bruns, M (ed), *Nouveaux modes de detection et de prevention de la discrimination et accès au droit: action de groupe et discrimination systémique, algorithmes et préjugés, réseaux sociaux et harcèlement*, 119

512 See Martini, M (2017), ‘Algorithmen als Herausforderung für die Rechtsordnung’ (Algorithms as a challenge for the legal system) 21 *JuristenZeitung* 1017, 1024.

513 Ibid.

514 Ibid.

515 Ibid.

516 Ibid.

517 De Barros Fritz, R and Von Lewinsky, K (2018), ‘Arbeitgeberhaftung nach dem AGG infolge des Einsatzes von Algorithmen bei Personalentscheidungen’ (Employer liability due to the use of algorithms in personnel decisions) *Neue Zeitschrift für Arbeitsrecht* 620.

518 Cavallaro, MC and Smorto, G (2019), ‘Decisione pubblica e responsabilità dell’amministrazione nella società dell’algoritmo’ (Public decision and responsibility of the administration in the algorithm society), *Federalismi.it* (4 September 2019) available at: https://www.federalismi.it/nv14/articolo-documento.cfm?Artid=40182&content=Decisione%2Bpubblica%2Be%2Bresponsabilità%2Bdell'amministrazione%2Bnella%2Bsocietà%2Bdell'algoritmo&content_author=%3Cb%3EMaria%2BCristina%2BCavallaro%2Be%2BGuido%2BSmorto%3C%2Fb%3E.

effects, but such problems of discrimination may be difficult to detect and challenge because of transparency problems and issues of responsibility and liability. The national reports show that, so far, this awareness has not resulted in strong efforts to introduce legislation to counter such problems. In none of the countries studied has new equality or non-discrimination legislation been adopted, and existing legislation has not been amended to deal with the challenges of algorithmic decision making. Nevertheless, the national experts for **Denmark, Germany, Greece, Malta and Norway** have suggested that the dynamics of the policy and public debates (to be discussed in section 3.3.1) are such that some legislative proposals in this field can be expected in the future.⁵¹⁹ More often, however, existing legislation in a variety of fields has been mentioned as relevant to help address certain aspects of algorithmic discrimination.

Non-discrimination and equal treatment legislation

In by far the majority of countries, existing non-discrimination legislation is relied on to deal with any issues of algorithmic discrimination that might occur. It has been reported that, in many countries, gender equality and non-discrimination legislation applies to algorithmic discrimination in most relevant spheres of life, for example in **Belgium, Bulgaria, Czechia, Denmark, France, Greece, Latvia, Malta** (under its pending equality bill), the **Netherlands, Norway, Portugal, Romania, Slovakia, Sweden** and the **United Kingdom**. The gender equality legislation in these countries not only covers employment-related algorithmic decisions, but mostly also (potential) algorithmic discrimination that relates to the provision of goods and services, the media, advertising and education.⁵²⁰ Sometimes the applicability of such general legislation to matters of algorithmic decision making is explicitly addressed. In **Sweden**, for example, the Equality Ombudsman has emphasised that the non-discrimination legislation also applies to algorithmic discrimination.⁵²¹ In **France**, the Defender of Rights⁵²² has clarified that his mandates – and by extension the national equality laws – also apply to issues of algorithmic discrimination.

Technology-specific legislation, sectoral legislation and general criminal and civil law provisions

In several countries, notably **Bulgaria, France and Malta**, legislation has been adopted that is considered to be potentially relevant to specific algorithmic applications, such as legislation on electronic communications or on detecting hate speech and cyber-harassment on social media platforms.⁵²³ Such legislation may not focus on equality and non-discrimination, and may not directly aim to provide protection against discrimination, but national experts have noted that such legislation can generally help to ensure that AI and algorithms are developed and used in line with ethical principles. According to national experts, it also may guarantee the presence of specific mechanisms for monitoring and auditing.

519 For example, in **Germany**, the Data Ethics Commission has proposed tightening the existing regulatory framework, for instance in relation to profiling and scoring: see Data Ethics Commission (2019), 'Expertise', available at: <https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenethikkommission/arbeitsergebnisse-der-dek/arbeitsergebnisse-der-dek-node.html>.

520 In some countries, such as **Latvia and Poland**, the applicability of the national legislation transposing the Goods and Services Directive is not entirely clear. In **Croatia**, the gender equality legislation does generally apply to media, advertising and education, but an important exception is made for the social media and networks.

521 Equality Ombudsman (2019), *Promemoria om automatiserad databehandling med algoritmer och risker för diskriminering: inom rekrytering och kreditgivning* (Memorandum on automated data processing with algorithms and risks of discrimination: in recruitment and lending) (Registration No 2018/387, Document 37, 4 October 2019) 44.

522 Defender of Rights and CNIL (National Data Protection Authority of France) (2020), *Algorithmes: prévenir l'automatisation des discriminations* (Algorithms: preventing automated discrimination) joint report, 7, available at: www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-en-num-16.07.20.pdf.

523 **Bulgaria**: Law on Electronic communications / Закон за електронните съобщения/-S.G. No. 41/ 2007, last amended on 5 June 2020: <https://www.lex.bg/laws/ldoc/2135553187>; **France**: *Loi du 24 juin 2020 visant à lutter contre les contenus haineux sur internet* – Bill aimed at fighting against hate contents on internet] which placed an obligation on online platforms and search engines to remove 'manifestly illicit contents' such as hate speech or racist abuse or cyber-harassment within 24 hours adopted by the National Assembly 13 May 2020, but cancelled in part by the Constitutional Council as an infringement of freedom of expression considering the difficulty to monitor all content closely, Constitutional Council, Décision No. 2020-801 DC of 18 June 2020, available at: www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm; **Malta**: legislation available at: <https://legislation.mt/eli/cap/9/eng/pdf>.

In some other countries, there is sectoral legislation in place that addresses specific aspects of the use of algorithms for decision making, e.g. in banking or healthcare, which might be relevant to gender equality and non-discrimination. For example, in **Poland**, a 'right to clarify' has been formulated in banking legislation, which is hoped to increase the transparency of creditworthiness decisions and potentially could be helpful in identifying and demonstrating algorithmic discrimination in such decisions.⁵²⁴ **France** has a bill pending on bioethics containing a duty to inform patients about the use of algorithms, which similarly could help individuals to find out about potential forms of discrimination.⁵²⁵ For **Germany**, it has been suggested that competition and consumer law might help to address some specific algorithmic challenges, for example by allowing consumer organisations to bring class actions and enforce the law without an identifiable victim.⁵²⁶

Finally, national experts have pointed out that general criminal law provisions (as reported in respect of **Croatia**,⁵²⁷ **Greece**,⁵²⁸ **Malta**,⁵²⁹ and **Slovakia**⁵³⁰) or sometimes general provisions of civil law (e.g. in **Poland**)⁵³¹ can be relied on to fight harassment, hate speech and incitement to violence that is enabled or exacerbated by the use of algorithms.

Data protection legislation

As discussed in section 3.3.1.1, national legislation to execute the EU's General Data Protection Regulation (GDPR), is generally considered to constitute a valuable addition to non-discrimination and equal treatment law. The expert for **France** has pointed out that non-discrimination law provides a grid of analysis as regards the use of certain data that also could be relevant for data protection, while, in turn, data protection tools impose useful duties of compliance that can also help to prevent and redress inequality and discrimination.⁵³² For example, the national expert mentions that Article 13 GDPR imposes the obligation to provide any useful information on automatic decision-making processes, which could, by analogy, be helpful in relation to the enforcement of equality and non-discrimination obligations.

524 Act of 21 February 2019 amending certain acts in connection with ensuring the application of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on free flow of such data and repealing Directive 95/46 / EC (General Data Protection Regulation) (Journal of Laws, 730).

525 The bill on bioethics (Text No. 686 of the General Assembly on August 3 2020 (Article 11)) was sent to the Senate (second reading) and is now pending in front of the Mixed Commission in charge of drafting a version which integrates the amendments of the Senate and General Assembly, see: www.senat.fr/leg/pjl19-686.html; the last version of the bill is available at: <https://www.senat.fr/leg/pjl19-686.html>.

526 See Martini, M (2017), 'Algorithmen als Herausforderung für die Rechtsordnung' (Algorithms as a challenge for the legal system) *JuristenZeitung* (21) 1017.

527 Article 325 (1) – (3) Criminal Code (*Kazneni zakon*), Official Gazette *Narodne novine* Nos 28/2013, 33/2015, 82/2015, 82/2016, 67/2018 and 126/2019.

528 Harassment (together with sexual harassment) is explicitly prohibited in national legislation: i) in the field of gender equality by Article 3(2)(a) of Act 3896/2010 transposing Directive 2006/54/EC; Article 5 of Act 3769/2009 transposing Directive 2004/113/EC; and Article 4(2) of Act 4097/2012 transposing Directive 2010/41/EU. Article 2(c) of Act 3896/2010 transposing Directive 2006/54/EC; Article 2(c) of Act 3769/2009 transposing Directive 2004/113/EC; and Article 3(c) of Act 4097/2012 transposing Directive 2010/41/EU, copying the definition from the Directives and therefore complying with the Directives; Article 2(10) Act 4604/2019, OJ A 50/26.03.2019 repeats the definition of 'harassment' contained in Article 2(c) of Act 3896/2010, adding to sex the grounds of sexual orientation and gender identity; ii) in the field of anti-discrimination law on the other grounds by Article 2(2c) Act 4443/2016, OJ A 232/09.12.2016.

529 <https://legislation.mt/eli/cap/9/eng/pdf>.

530 Slovak Republic, Act No. 300/2005, Criminal Code – namely: Stalking (Article 360a), Extortion (Article 189), Duress (Article 192), Sexual Exploitation (Articles 201, 201a, 201b), Defamation (Article 373), Harm Done to Rights of Another (Articles 375, 376), Manufacturing of Child Pornography (Article 368), Dissemination of Child Pornography (Article 369), Possession of Child Pornography and Participation in Child Pornographic Performance, Corrupting Morals (Articles 371, 372), Corrupting Morals of Youth (Article 211), Establishment, Support and Promotion of Movements Directed at the Suppression of Fundamental Rights and Freedoms (Article 421), Expression of Sympathy for Movements Directed at the Suppression of Fundamental Rights and Freedoms (Article 422), Production, Distribution, Possession of Extremist Materials (Articles 422a, 422b, 422c), Denial and Approval of the Holocaust, the Crimes of Political Regimes and Crimes against Humanity (Article 422d), Defamation of Nation, Race and Conviction (Article 423), or Incitement to National, Racial and Ethnic Hatred (Article 424).

531 Article 23 of the Civil Code, Act of 23 April 1964, consolidated text, JoL of 2018, item 1025, as amended.

532 Similarly, the expert for **Poland** has remarked that 'the provisions on the protection of personal data are seen as natural and obvious support in attempts to regulate the issue of AI and the fight against injustices generated by ADM, including algorithmic discrimination'.

Similarly, in **Germany**, the transparency provisions in Articles 13–15 and the obligation to carry out data protection impact assessments of Article 35, as well as the associated consultation obligation under Article 36 GDPR and the documentation obligations under Article 30 GDPR, are considered to play an important supporting role where fighting algorithmic discrimination is concerned. In some states, such as **Sweden**, additional rights have been granted in this respect on the national level. In particular, as the national expert has explained, anyone who is subject to the prohibition of discrimination should be provided with information if the Equality Ombudsman so requests.⁵³³ The individual could then use such information to identify and challenge algorithmic discrimination. Indeed, several experts (in particular for **Denmark, France, Ireland, the Netherlands and Sweden**) have pointed out that knowing that and how one's data is used may enable individuals to know whether certain potentially problematic factors (such as protected grounds) are taken into account in the processing of their data, to detect potential (direct) discrimination. In addition, this obligation is considered to have the added benefit that it helps to reduce the opacity of the workings of algorithms, which is a major problem identified in relation to algorithmic decision making.

The provisions of the GDPR have also been mentioned as useful in establishing responsibility for algorithmic decisions and promoting awareness and accountability (**Germany, Italy, Romania and Spain**). In **Spain**, according to the Spanish Agency for Data Protection, it is compulsory to run some kind of risk-based analysis, specifically concerning the risks to rights and freedoms regarding the treatment of personal data and the preparation of profiles. In determining the level of risk of an AI-based treatment, the agency's guidelines highlight the risks of bias in automatised decision making and algorithmic discrimination.⁵³⁴ The Spanish Agency for Data Protection has also emphasised that whoever adopts a decision that includes AI cannot argue that they have insufficient information or technical knowledge to avoid responsibility when auditing or deciding on the adequacy of their system to meet the legislative requirements.⁵³⁵ This means that, according to the agency, the responsibility cannot be allocated to the developer or provider of the tool, and even less, to the AI system itself. It is the client who is responsible for testing and auditing the treatment that is produced by the AI-based system.

Furthermore, the experts for **Denmark, Estonia, France, Germany, Greece, Iceland, Liechtenstein, the Netherlands and Spain** have pointed out that Article 22 of the GDPR prohibits automated decision making and profiling. Since such uses of data are very closely associated to algorithmic decision making and are generally regarded as strongly related to problems of discrimination, the experts regard GDPR as a potentially useful tool in fighting algorithmic discrimination. It should be added as a nuance here, however, that fully automated decision making using algorithms is currently rather uncommon, and Article 22 GDPR applies neither to semi-automated decision making, nor to the use of algorithms as an aid to decision making.

Experts for **Iceland, Italy, Liechtenstein, Norway, Poland, Slovakia and Spain** have noted that Article 5 of the GDPR contains limitations on the use of 'personal data' in decision making. Although Article 5 does not expressly refer to gender, it is considered that the concept is likely to cover gender or gender-related data.⁵³⁶ Interpreted in that manner, the provision could be used to help prevent the use of biased or sensitive data in algorithm-driven decision making.

533 Chapter 4, Section 3 of the Discrimination Act (2008:567). See also Equality Ombudsman (2019), *Promemoria om automatiserad databehandling med algoritmer och risker för diskriminering: inom rekrytering och kreditgivning* (Memorandum on automated data processing with algorithms and risks of discrimination: in recruitment and lending) (Registration No 2018/387, Document 37, 4 October 2019) 44.

534 Agencia Española de Protección de Datos, available at: <https://www.aepd.es/es>.

535 Spanish Agency on Data Protection (2020), *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial* (Adaptation to the GDPR of treatments that incorporate artificial intelligence) available at: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf.

536 As a nuance to this position, it can be remarked that there are important limitations to this, e.g. the list of 'special categories of personal data' in Article 9 does not include gender/sex or disability or age.

Finally, it has been remarked that the interface between equality and non-discrimination law and data protection law may mean that not only equality bodies, but also data protection authorities are competent to deal with issues of equality and non-discrimination. For **Finland**, for example, it has been explained that the provisions on automated decision making and data impact assessments have the consequence that the Data Protection Ombudsman is competent to deal with specific issues of algorithmic discrimination.⁵³⁷ Similarly, in the **Netherlands**, the Data Protection Authority recently published a report in which it found that the Netherlands Tax and Customs Administration had taken nationality into account in carrying out a fraud-detection policy.⁵³⁸ According to the Data Protection Authority, this not only constituted prohibited processing of personal data, but also resulted in discrimination. It was able to address this particular aspect of discrimination by interpreting the relevant GDPR provisions (especially Article 5) in the light of international treaty provisions on non-discrimination, such as Article 26 of the International Covenant on Civil and Political Rights and Article 1 of Protocol 12 to the European Convention on Human Rights. The coexistence and overlapping competence of such bodies may necessitate stronger cooperation and harmonisation.

3.4.1.2 Problems and gaps in national equal treatment and non-discrimination legislation

It has been shown in section 3.4.1.1 that hardly any specific legislative efforts have been made on the national level to address the specific discrimination challenges related to algorithmic decision making. This raises the question of whether the existing legislation is generally perceived to offer sufficient protection against cases of algorithmic discrimination.

Overall, insofar as gender equality and non-discrimination is concerned, most national experts think the national legislation has a sufficiently wide material scope to cover most examples of algorithmic discrimination. This is different for countries where the national legislation transposing the Goods and Services Directive has made an exception for media, advertising and education, as is the case in **Austria, Cyprus, Estonia, Finland, Germany, Greece, Italy, Liechtenstein, Norway, Poland, Portugal, and Romania**.⁵³⁹ The exception has been said to raise problems in relation to algorithmic decision making in particular in **Cyprus, Germany, Greece and Italy**, where the experts have emphasised that algorithmic decisions are particularly prevalent in these sectors. Indeed, this perceived prevalence is confirmed by the examples of targeted advertisements on social media and the use of algorithms in school selection procedures provided in section 3.1. Thus, in countries where the exception has been used without there being any compensation for this in other legislation or soft-law instruments, the gender equality legislation is regarded as displaying serious gaps in the protection against algorithmic discrimination.

In addition, other deficiencies have been identified in national legislative protection against algorithmic discrimination. For a number of countries, some concerns arise in relation to the fact that the material scope of equality and non-discrimination legislation is limited for grounds other than gender. In **Denmark**, for example, it is considered problematic that gender-based discrimination is offered far stronger and wider protection than discrimination based on, for example, religion and sexual orientation. National experts have made similar comments for **Greece, Poland and Romania**. In respect of **Belgium and Italy**, moreover, it has been pointed out that a relatively narrow definition has been given to the notion of gender, which has the result that gender equality provisions, regardless of their generally wide material scope, do not offer any protection to intersex or transgender persons. Finally, a specific exemption in the legislation transposing the Goods and Services Directive in **Germany** (Section 3(4) of the General Equal

537 On the competence of the Data Protection Ombudsman, see <https://tietosuojafi/en/automated-decision-making-and-profiling>.

538 Autoriteit Persoonsgegevens (Tax and Customs Authority) (2020), *Belastingdienst/Toeslagen: De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag* (The processing of the nationality of recipients of child care benefits) (Onderzoeksrapport z2018-22445) available at: www.autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf.

539 See EELN (2019), *A Comparative Analysis of Gender Equality Law in Europe* (Brussels, European Commission) available at: www.equalitylaw.eu/publications/comparative-analyses, 124. This report also shows that in several other countries, in-between or unclear positions have been taken.

Treatment Act) means that sexual harassment is not considered discrimination in access to goods and services provided under civil law contracts. For the provision of goods and services under public law, harassment and sexual harassment are not considered discrimination in this field and the special rules on support by anti-discrimination organisations do not apply.

In section 2.1.1, it was explained that EU legislation on pregnancy and maternity protection and on work-life balance could offer additional safeguards in relation to algorithmic decision making, including in light of its rather wide material scope. On the level of the Member States, however, this legislation is hardly seen to offer protection against algorithmic discrimination. A majority of national experts deemed it unclear, at the moment, how the legislation could materially contribute to preventing or redressing algorithmic discrimination given the early stages of existing discussions and the still limited awareness of the issue. Nevertheless, for **France**, it was mentioned that this legislation could help by prohibiting the reference to pregnancy, maternity and parenting in the conception of algorithms that target certain profiles for pricing products or services. In addition, some national experts (e.g. for **Germany, Italy, Portugal** and **Spain**) explained that the legislation might be of assistance indirectly, since it could contribute to reducing gender-based stereotypes. In the long run, it could thus help to ‘clean’ data and have a positive impact on the correlations found by algorithms.

In some countries, specifically **Belgium, Bulgaria, Czechia, Germany, Iceland, Liechtenstein, Lithuania, Norway** and **Portugal**, national experts regret that there are no technology-specific provisions on discrimination or provisions related to algorithms, because they find that the current legislation is sometimes difficult to apply, does not provide for sufficient legal certainty and clarity, and sometimes does not sufficiently address the use of algorithms by private parties. In **Germany**, for example, it has been questioned whether the equal treatment legislation, particularly in relation to goods and services, is sufficiently capable of dealing with problems of automated decision making, since such decisions may result in ‘individual’ rather than ‘mass contracts’, which are not covered by the applicable legislation.⁵⁴⁰ In addition, for **Malta** and **Portugal**, national experts have observed problems with respect to the lack of tangible legislative provisions on transparency in algorithmic decision-making processes, which may make it difficult to effectively contest a case of algorithmic discrimination.

One specific issue that can arise in relation to algorithmic discrimination is that of intersectionality and multiple discrimination. It was explained in section 1.4.3 that algorithms are very good at identifying a wide range of relevant datapoints, proxies and factors that are relevant in detecting useable patterns and profiles. Although it may be possible to reduce such factors and datapoints to one particular protected ground (as explained in section 2.2), usually, the actual basis of a decision will be granular and diverse. At best, the factors that are decisive for the output suggested by the algorithm are a combination of protected grounds or proxies that are very closely correlated to the protected grounds. It has been mentioned by national experts that the tendency of algorithmic decision making to focus on combinations of or intersections between (protected) grounds or combinations of characteristics can be difficult to deal with. It can be most easily dealt with by equality and non-discrimination laws that either do not have an exhaustive list of grounds, as is the case in **Finland**, or specifically address multiple or intersectional discrimination, as has been reported for **Belgium, Czechia, Germany, Malta** (in a pending bill) and **Norway**. Even then, however, as was remarked for **Germany**, the case law available may be too limited to confirm whether the relevant provision would indeed cover the specific issues raised by algorithmic intersectional discrimination.

Dealing with intersectional discrimination is much more difficult in states where individual cases have to be brought for each case of discrimination (**Ireland, Spain**), where cases have to be reduced to one particular protected ground (**Denmark**) or dealt with under the specific laws and exceptions related to the different individual grounds (**Netherlands**), or where the ‘strongest’ ground in a case is chosen

540 Section 19(1)(1) of the General Equal Treatment Act; see on this point also Berghahn, S, Klapp, M and Tischbirek, A (2016), *Evaluation des Allgemeinen Gleichbehandlungsgesetzes* (Evaluation of the General Equal Treatment Act) (ADS, Berlin) 116.

(Italy). Similarly, dealing with the difficulties related to proxy discrimination may be difficult in countries for which the national experts have reported that only the concept of multiple or ‘double’ discrimination is recognised, but not that of intersectionality (**Croatia, Greece, Italy, Romania and Slovakia**). Very little protection against the particularities of algorithmic discrimination is further offered in countries where not even the notion of multiple discrimination is recognised in the legislation or in legal practice, as national experts have noted for **Bulgaria, Cyprus, Hungary** (although the equality body has reportedly acknowledged the concept of intersectionality), **Latvia, Portugal, Spain, Sweden** and the **United Kingdom**.

Finally, it has been noted in sections 3.3.1.2 and 3.4.1.1 that – to some degree – the limitations of equality and non-discrimination legislation in offering protection against algorithmic discrimination can be compensated for by other legislation, in particular legislation on the protection of personal data. However, it has been remarked by the expert for **Belgium** that data protection legislation does not address all matters relevant to algorithmic discrimination, such as matters of bias in the data used to train an algorithm. In addition, it was shown that there may be a clear overlap between data protection legislation and non-discrimination law on the national level. Although this may lead to valuable cross-fertilisation, it can also create a need for enhanced cooperation and harmonisation between equality bodies and data protection authorities.⁵⁴¹

3.4.2 (Semi-)judicial application and enforcement of legislation

3.4.2.1 Relevant judgments and decisions by semi-judicial bodies in the European countries

In by far the majority of countries, no case law is yet available on specific cases of gender inequality or discrimination caused by algorithms. To the extent that cases on algorithms are brought before the national courts, they usually relate to specific aspects of algorithmic decision making, such as transparency and data protection. For example, in the **Netherlands**, the highest administrative court has found that there is a general obligation for public authorities to ensure explainability, transparency and accessibility of algorithms in order for individuals to understand how they have been affected by an algorithm and to enable them to effectively contest that algorithm before a court.⁵⁴² This position has been embraced by the Supreme Court of the Netherlands.⁵⁴³ In another recent Dutch judgment, the legislation that allowed the use of a predictive profiling algorithm in detecting social security fraud (SyRI) has been found to be incompatible with the general right to privacy, since individuals were given too little information about the way in which the algorithm operated and used their data.⁵⁴⁴ While discrimination complaints were made by the parties in the SyRI case, they did not play any significant role in the judgment, although the court in its reasoning showed explicit awareness of the potential risks of biased and stereotype-based decision making involved in the use of the particular profiling algorithm involved.

Similarly, in **Poland**, cases have been brought on STIR, a profiling system for categorising unemployed people⁵⁴⁵ and on an algorithm introduced to randomly assign court cases.⁵⁴⁶ In both cases, however, issues of algorithmic discrimination were not central to the case, and the judgments concentrate on matters such as algorithmic accuracy, information on source codes and transparency. **France** has also seen some

541 In **France**, such shared responsibility is already seen on the national level, and there is already a degree of collaboration; see e.g. the joint report by the Defender of Rights and CNIL (2020), *Algorithmes: prévenir l'automatisation des discriminations* (Algorithms: preventing automated discrimination), 7 available at: www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-en-num-16.07.20.pdf.

542 ABRvS (Judicial Division of the Council of State) 17 May 2017, ECLI:NL:RVS:2017:1259 (Aerius I); ABRvS 18 July 2018, ECLI:NL:RVS:2018:2454 (Aerius II).

543 HR (Supreme Court) 17 August 2018, ECLI:NL:HR:2018:1316 (WOZ).

544 Rechtbank (District Court) Den Haag 5 February 2020, ECLI:NL:RBDHA:2020:865 (SyRI); see also section 3.1.2.4.

545 See further on STIR section 3.1.2.4. See also Judgment of the District Administrative Court in Warsaw, case no III SA / Wa 2057/18; in the judgment, the court did not seem to pronounce explicitly on the algorithm used.

546 District Court Warsaw 5 September 2018, Case no II SAB / Wa 61/18 (finding that the source code did not need to be disclosed); a cassation appeal is pending.

constitutional litigation on the use of algorithms in relation to its new legislation on flagging hate speech on (social media) platforms.⁵⁴⁷ Again, however, this litigation did not concentrate on equality issues, but rather dealt with the difficulties involved in closely monitoring the content published on social media.

Further, some examples can be found of cases on discrimination by algorithms that have been decided or are currently pending before the national courts, equality bodies or other semi-judicial bodies. The subject matters of these cases are diverse: in **Denmark**, the Danish Human Rights Institute has filed a complaint with the Equal Treatment Board on targeted advertising;⁵⁴⁸ in **Finland** and **Sweden**, cases have been brought about automated decisions in relation to credit ratings, bank loans and credit purchases;⁵⁴⁹ in the **United Kingdom**, a predictive profiling algorithm using automatic facial recognition by South Wales Police has been contested;⁵⁵⁰ in **France**, there has been intensive litigation on the algorithmic selection process used to allocate places on higher education courses;⁵⁵¹ and finally, cases related to the use of algorithms in making employment-related decisions (for instance in recruitment or job transfers) and employment-related discrimination by platforms have been considered by courts in **Italy**⁵⁵² and the **United Kingdom**.⁵⁵³

From the judgments and decisions in both equality and non-equality related cases of algorithmic decision making, it can be presumed that some courts and equality bodies have found it problematic if decisions were automatically based on the outcome of the algorithm and without there being a 'human in the loop' to check and, if need be, correct the decision (e.g. in **Italy**⁵⁵⁴ and **Poland**⁵⁵⁵). In other cases, algorithms have been held to generate factually inaccurate outcomes, which showed their overall deficiency, as was shown in a **Polish** case on Amazon's automated dismissal policy.⁵⁵⁶ Furthermore, the amount of information to be disclosed about the functioning of the algorithm is often a contested issue. As mentioned above, courts in the **Netherlands** have accepted that explainability and provision of information are extremely important in cases where algorithms are used as a basis for decision making by public authorities.⁵⁵⁷ Similarly, the Constitutional Court of **Slovakia** has decided that a decision to introduce a SARS-COV-19 track-and-trace app was not sufficiently clear about how the data would be processed by that app and what guarantees would be offered in relation to their use and storage.⁵⁵⁸ By contrast, the Council of State in **France** did not see a need to publish details on the functioning of the algorithm that supported decisions relating to the admission of students to institutions of higher education,⁵⁵⁹ although this was

547 Bill to fight against hate speech on internet, platforms and search engines in 24 hours adopted by the National Assembly 13 May 2020 but cancelled in most part by the Constitutional Council as an infringement of freedom of expression considering the difficulty to monitor all content closely, Constitutional Council, Décision No. 2020-801 DC du 18 June 2020, <https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm>.

548 See <https://menneskeret.dk/nyheder/instituttet-klager-ligebehandlingsnaevnet-to-virksomheders-maalrettet-annoncering>.

549 Case 216/2017, of 21.3.2018, https://www.yvtltk.fi/material/attachments/ytalk/tapausselosteet/2SVkNzOWF/YVTltk-tapausseloste-21.3.2018-luotto-moniperusteinen_syrjinta-S_L.pdf; Sweden: decision by the Equality Ombudsman, GRA 2017/80.

550 *R (Bridges) v CCSWP and SSHD* [2019] EWHC 2341 (Admin).

551 Administrative Guideline 24 April 2017 cancelled by the Conseil d'Etat (*Circulaire 24 avril 2017 annulé par le Conseil d'Etat*), 22 December 2017 Association SOS Education, Promotion et défense des étudiants et Droits des Lycéens Nos. 410561, 410641, 411913; Council of State No. 427916, 12 June 2019, <https://www.conseil-etat.fr/ressources/decisions-contentieuses/dernieres-decisions-importantes/conseil-d-etat-12-juin-2019-universite-des-antilles>; Constitutional Council, Decision No. 2020-834 QPC, 3 April 2020.

552 TAR Lazio decisions No. 10964/2019, No. 9924/2018 and No. 6606/2019; Consiglio di Stato decisions: No. 881/2020; Nos. 8472, 8473, 8474/2019; No. 2936/2019, No. 2270/2019.

553 *GLS v Brookes* [2017] IRLR 780; *Gibbs v Westcroft* [2014] 12 WLUK 110.

554 Consiglio di Stato decisions: No. 881/2020; Nos. 8472, 8473, 8474/2019; No. 2936/2019, No. 2270/2019.

555 Judgment of 29 September 2016, reference number act II SA / Po 395/16.

556 District Court of Poznań-Grunwald and Jeżyce, sitting in Poznań, partial judgment of 5 July 2018, case no. VP 695/16; see also <https://www.wiadomoscihandlowe.pl/artykuly/w-sadach-pracy-toczy-sie-12-postepowan-przeciw-ama,48710>. The example was discussed in section 3.1.3.1.

557 ABRvS (Judicial Division of the Council of State) 17 May 2017, ECLI:NL:RVS:2017:1259 (Aerius I); ABRvS 18 July 2018, ECLI:NL:RVS:2018:2454 (Aerius II); HR (Supreme Court) 17 August 2018, ECLI:NL:HR:2018:1316 (WOZ).

558 See www.ssn.sk/22860/ustavny-sud-slovenskej-republiky-informuje-o-dovodoch-rozhodnutia-vo-veci-vedenej-pod-sp-zn-pl-us-13-2020/.

559 Council of State No. 427916, 12 June 2019, <https://www.conseil-etat.fr/ressources/decisions-contentieuses/dernieres-decisions-importantes/conseil-d-etat-12-juin-2019-universite-des-antilles>.

later slightly nuanced by the Constitutional Council.⁵⁶⁰ Similarly, in **Poland**, it has been decided that the data used to feed the algorithm used by the random assignment system for court cases did not constitute public information and therefore could be kept secret.⁵⁶¹ In **Italy**, the Council of State has found that the (private) owner of an algorithm has a right to intellectual property and business secrets that is of equal weight to the interest of the disadvantaged person in knowing more about how the algorithm works.⁵⁶²

It can be concluded from the above that, thus far, there has been limited litigation in the field of algorithmic discrimination. Insofar as relevant judgments and decisions relate to other matters that are relevant to algorithmic decision making, such as overall transparency, accuracy, reliability and secrecy, they show that different courts in different countries may hold very different views.

3.4.2.2 Problems and gaps related to the (semi-)judicial application and enforcement of national equal treatment and non-discrimination legislation

In light of the challenges set by algorithm-based decision making, some particular problems and risks have been highlighted by national experts that could be relevant to the (semi-)judicial application and enforcement of national equal treatment and non-discrimination legislation. One of these relates to more general problems experienced in the enforcement of gender equality and non-discrimination legislation. In **Romania**, for example, there is reportedly little knowledge of non-discrimination legislation in courts in general, and it is expected that this will be no different for algorithmic discrimination. Similarly, national experts have mentioned that little protection against discrimination can be expected to be provided in **Greece, Italy** and **Spain**, since in these countries, hardly any discrimination cases are brought before the national courts in general and, if they are, there may be long delays and, in the end, such proceedings do not offer redress to the victim.

In some other countries, it is expected that algorithmic discrimination may raise specific issues in this regard. For example, in **Croatia**, the national expert expressed her concern that in cases on discrimination by algorithmic applications, expert witnesses are needed, which would be expensive and would make the effectiveness of judicial proceedings heavily dependent on their availability and training. For **Poland**, another problem was highlighted, which is that some technical issues, such as profiling, have been held not to be subject to appeal.⁵⁶³

Even if effective access to a court or similar (non-judicial) body is provided, the national experts have mentioned some specific problems in relation to enforcing non-discrimination legislation and offering redress to victims. The most important difficulty is related to the detection of discrimination in cases of algorithmic decision making, as was also discussed in section 3.2.4. Many experts have noted that as a result of the lack of transparency and the complexity of algorithms, it will be extremely hard to know whether their use results in a form of unequal treatment and discrimination that is directly related to gender or other protected grounds, or possibly a kind of unprotected inbetween ground. The problem of showing direct discrimination by an algorithmic application is further exacerbated if discriminatory intent

560 Constitutional Council, Decision No. 2020-834 QPC, 3 April 2020.

561 District Administrative Court of Warsaw 5 September 2018, case no II SAB / Wa 61/18; see also Supreme Administrative Court's judgment of 27 February 2014, case No. I OSK 2014/13.

562 Consiglio di Stato decision No. 30/2020. See also <https://brevettinews.it/internet-domini/titolare-dei-diritti-algoritmo-puo-oppor-si-all'accesso-agli-atti/>.

563 As the national expert for **Poland** explained, in 2016, an unemployment profiling system was appealed by the Ombudsman to the Constitutional Tribunal. In his application, the Ombudsman accused the provisions of being unconstitutional, due to the lack of an ability to appeal against the award of the profile and the fact that the scope of data used for profiling was not defined in a legal regulation in the order of statutory legislation. For more information, see Niklas, J., Sztandar-Sztanderska, K and Szymielewicz, K (2015), *Profiling the unemployed in Poland: social and political implications of algorithmic decision making* (Fundacja Panoptykon, Warsaw 2015) available at: https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf, 22 and 23. For the judgment of the Polish Constitutional Tribunal, see Wyrok Trybunału Konstytucyjnego z 6 czerwca 2018 r., sygn. K 53/16, Opublikowano: OTK-A 2018/38, Dz.U.2018/1149 (Case K 53/16, Judgment of the Constitutional Tribunal of 6 June 2018, Journal of Law 2018/1149) available at: <https://sip.lex.pl/orzeczenia-i-pisma-urzedowe/orzeczenia-sadow/k-53-16-wyrok-trybunalu-konstytucyjnego-522591612>.

must be shown (as was mentioned for **Germany**),⁵⁶⁴ since the expectation often will be that the algorithm is 'neutral' (as was mentioned in respect of **France**).

Related to the problem of detection of algorithmic discrimination, it has been pointed out by the national experts for **France**, **Germany** and **Poland** that one of the inherent problems of the current equality and non-discrimination legislation is that it focuses so strongly on individual cases, while many instances of algorithmic discrimination derive from more structural issues. Consequently, they can only be or are better addressed on a collective level than by means of individual court cases or complaints. In this regard, it has been suggested by the national expert for **Germany** that class action would be the only feasible way to bring a complaint about algorithmic discrimination before a court.

In some cases, and to some degree, the problems of detecting discrimination can be addressed by relying on the concept of indirect discrimination. However, this concept is not always easily recognised on the national level, and even if it is, it may be difficult to establish a *prima facie* case of indirect discrimination due to the equality and non-discrimination practice existing in a certain country, combined with rules of evidence and the burden of proof. In some countries, strong evidence is needed to present a case of indirect discrimination to a national court, which might be difficult to provide anyway (e.g. in **Germany**).⁵⁶⁵ Offering such proof may be even harder in algorithmic discrimination cases, as was mentioned in respect of **Croatia**, due to the problems of transparency and secrecy discussed in section 3.2.3. Even if the rules of evidence are more lenient, as is the case in **Denmark** (where the burden of proof shifts once a correlation with a protected ground can be shown), **Germany** (where the burden of proof shifts when the facts presented – from an objective point of view and with a predominant probability – suggest that the disadvantage was at least also due to one of the forbidden grounds of discrimination)⁵⁶⁶ or **Ireland** (where the concept of indirect discrimination is purported to offer many opportunities to contest algorithmic discrimination), courts may not always be eager to accept statistical analyses or circumstantial evidence as sufficient proof of indirect algorithmic discrimination.⁵⁶⁷ In addition, in more lenient systems, the obscurity of algorithmic decision making processes can make it difficult to obtain the necessary information to show a *prima facie* case of indirect discrimination, as has been emphasised by the national experts for **France**, **Malta** and the **United Kingdom**. For some countries, such as **Denmark**, this can be solved to some degree by obliging public authorities to inform the recipient of a non-favourable administrative decision of the grounds for the decision, but the national expert has also pointed out that this obligation does not apply to private actors. Consequently, this obligation offers only a partial solution. Similarly, in other countries where such a right to information exists, such as **Sweden**, it has been noted that there are so many exceptions to this right that indirect discrimination is still difficult to prove.

The problem of detecting and proving (indirect) discrimination can be further exacerbated by intellectual property issues and protection of state or trade secrets, which may be a reason for non-disclosure of relevant information on the workings of an algorithm. It has been pointed out that in **Italy**, for example, the right to information does not automatically outweigh the intellectual property rights of owners of an algorithmic application, which may make it even more difficult to show the indirectly discriminatory nature of such an algorithm.⁵⁶⁸

Finally, it can be derived from the discussion in sections 1.4.6, 3.2.5, 3.3.1.3 and 3.3.2.3 that one of the main legal challenges related to algorithmic decision making relates to responsibility. Because there

564 See e.g. Sacksofsky, U (2010), *Mittelbare Diskriminierung und das Allgemeine Gleichbehandlungsgesetz* (Indirect discrimination and the General Equal Treatment Act) (Berlin) 21ff.

565 Berghahn, S, Klapp, M and Tischbirek, A (2016), *Evaluation des Allgemeinen Gleichbehandlungsgesetzes* (Evaluation of the General Equal Treatment Act) (ADS, Berlin, 2016); on the division of the burden of proof, see also Federal Labour Court, judgment of 22 July 2010, 8 AZR 1012/08, overruling the State Labour Court of Berlin and Brandenburg, judgment of 26 November 2008, 15 Sa 517/08.

566 Section 22 of the General Equal Treatment Act.

567 As suggested for **Germany**; see e.g. Berghahn, S, Klapp, M and Tischbirek, A (2016), *Evaluation des Allgemeinen*.

568 See specifically Consiglio di Stato decision No. 30/2020; see also <https://brevettinews.it/internet-domini/titolare-dei-diritti-algoritmo-puo-opporsi-allaccesso-agli-atti/>.

are often many bodies involved in the development and use of algorithms, it may not be easy to know to whom to address a complaint or court case. This problem has been emphasised on the national level, notably for **Hungary, Italy, Norway, Poland, Romania, Slovenia** and **Spain**. As a consequence of this difficulty, it has been reported that it may be very difficult for individual victims of algorithmic discrimination to obtain effective redress.

3.5 Conclusion

All over Europe, national experts have reported a clear readiness to use algorithms to support or drive decision making in the public and private sector. The examples provided in this chapter show a diverse and wide range of uses as well as a variety of aims and objectives that are pursued by the use of algorithms. The examples also show that in the public sector, the use of algorithms is often still at an experimental or pilot stage, and it is often reported to be rather controversial. Nevertheless, a considerable number of examples of (projected) reliance on algorithms can be seen in public policy fields ranging from labour market policy and social welfare to the administration of justice, policing and fraud detection. In the private sector, the purposes of algorithmic applications are even more varied than in the public sector. Various examples have been given of algorithms being used in employment (in particular, in recruitment), in banking and insurance, and for the purposes of targeted advertising, pricing and retail.

It is not always clear what types of algorithms are used in these different examples, but in light of their uses and objectives, it can be assumed that these are mostly machine-learning algorithms, because of their capacity for profiling, detecting patterns, and predicting human behaviour. In both sectors, moreover, enabling technologies are used in combination with algorithms, as in the example of detecting human emotions through analysing facial expressions and traits and using them as a basis for profiling.

With respect to these (projected and currently used) algorithmic applications, this chapter has shown that the experts recognise nearly all the six main challenges identified in Chapter 1, with the exception of the scale and speed challenge:

1. The *'human factor' and the stereotyping and bias challenge* is considered to play an important role. Experts note that algorithms can easily cause direct and indirect discrimination, mostly as a result of inbuilt biases and stereotyping, and report that the use of algorithms in decision making can perpetuate existing conditions of structural discrimination and prejudice.
2. The *data challenge* mainly relates to the accuracy, quality and reliability of data. It is clear from this chapter that this is cause for concern. In many examples, no corrections are made and no compensation is provided to account for negative societal stereotypes that may have entered the datasets on which algorithms are trained. In addition, some examples have been given of algorithms that were based on clearly incorrect, inaccurate or biased data, but because of the lack of openness and transparency (see below), it is often difficult to find out about such problems and challenge them effectively.
3. The *proxies challenge* has been noted to be particularly present in relation to intersectional discrimination. Algorithms are seen to be very well capable of identifying factors and parameters that are highly individual in nature and may often be at the intersection of different protected characteristics. The experts recognise that equality and non-discrimination laws that are strongly 'ground based' may not be easily invoked to obtain redress if one particular ground is not the basis of a negative decision, but rather a proxy or a complex combination of grounds.
4. The *transparency and explainability challenge* is frequently mentioned in relation to nearly all examples of algorithmic applications in the European countries. According to many experts, the lack of clear and comprehensible information on the workings of algorithms results in a problematic lack of control over their outcomes. This is considered to be even worse in situations where the source code or other aspects of an algorithm are kept secret for operational reasons. In addition, the opacity of algorithmic applications is considered as creating significant difficulties in detecting

and identifying algorithmic discrimination as well as in challenging them effectively before equality bodies and courts.

5. In legal terms, many experts have highlighted the *responsibility challenge*. Even if a case of (indirect) algorithmic discrimination can be detected, it may be difficult to contest it because of the compound responsibility that arises from the involvement of many different individuals and organisations in the process of planning, developing and using an algorithm. It is then difficult to know whom to hold liable in court proceedings.

In addition to these algorithm-related challenges, it has been found that a 'gender digital gap' continues to exist in the European countries. The IT sector struggles to attract and retain a female workforce, and minority groups also remain underrepresented. Since IT specialists are often the ones involved in designing and developing algorithmic applications, this is problematic. A more diverse workforce would be beneficial to creating a stronger awareness of the risks and challenges of biased data and stereotyping that are inherent to algorithms.

On the national level, there appears to be considerable awareness of these risks and challenges, although there are clear differences between countries and actors. In about one third of countries, some public discussion on algorithmic discrimination can be noted as part of, often broader, public debates on AI and ethics or AI and fundamental rights. These debates are often fuelled by non-governmental organisations and other civil society organisations and the media. In another (slightly larger) set of countries, specific discussions on the impact of algorithms on equality rights have been much more limited, and debates have tended to focus on broader ethical or related legal questions such as transparency, data protection, privacy and the use of algorithms by public authorities. In the final group of countries, national experts indicate that the issue of algorithmic discrimination has not yet permeated the public sphere to any significant extent.

As for scholarship in the field, it is clear that the discriminatory impact of AI is an emerging topic of interest in all 31 countries covered in this report. The number of workshops organised, studies conducted and articles published on the topic of ethical AI and algorithms has mushroomed in the past few years. Remarkably, much of the academic work reported focuses on the US context or international regulatory or ethical initiatives (for example at the level of the Council of Europe) rather than on national contexts. Empirical data on, and scientific analysis of, the impact of algorithms on discrimination in national contexts remain scarce.

The various challenges and problems of algorithmic discrimination have drawn some, but no significant attention from national legislators. So far, none of the European countries has introduced specific legislation to deal with the risks of algorithmic discrimination. By far the majority of countries rely on existing legal instruments to tackle the risks and challenges, such as non-discrimination and equality legislation and data protection laws. In countries where general non-discrimination and equality legislation is in place, and the legislation has both a wide material and personal scope, this has been reported to be a relatively satisfactory response. Nevertheless, even then, and as is also further discussed below, such legislation does not help to overcome the transparency and responsibility challenges, or make it easier to detect and effectively challenge cases of algorithmic discrimination.

Significantly less relevant legal protection is offered in countries where the provision of goods and services, the media, advertising and education have been excluded from the substantive scope of protection of the Gender Goods and Services Directive. This is considered even more problematic because these are the fields in which algorithmic discrimination currently occurs most often. In countries where the exception has been used without there being any compensation for this in other legislation or soft-law instruments, the gender equality legislation is therefore regarded as showing serious gaps in the protection against algorithmic discrimination. Similarly, for a number of countries, it is mentioned with some concern that the material scope of equality and non-discrimination legislation is limited for grounds other than gender, resulting in an uneven and patchy protection against algorithmic discrimination.

It is further considered problematic that in a rather large number of countries, equality and non-discrimination legislation does not provide for any form of express protection against intersectional (or even multiple) discrimination. This means that the legislation is unable to cope with the particular challenges of proxy discrimination that algorithmic applications pose and that have been described in sections 1.4.3 and 2.2.

Several experts have remarked that data protection laws can offer additional protection against some of the problems and risks of algorithmic decision making and can fill some gaps. In particular, the requirement to provide access to data is reported to help increase algorithmic transparency, detect algorithmic discrimination, and identify the responsible persons or bodies. In addition, the requirement to conduct a data protection impact assessment and the obligation to have a data protection authority have been referred to as useful instruments to help avoid as well as monitor access to and use of sensitive data in algorithms. However, it also has been remarked that the coexistence and overlapping competence of equality bodies and data protection authorities may necessitate stronger cooperation and harmonisation. In addition, the GDPR has been criticised for its rather weak approach to discrimination.

Clarification of general equal treatment legislation and its application to specific cases of algorithmic discrimination is needed to better understand the current limitations and possibilities. As yet, however, the number of court cases that are expressly related to algorithmic discrimination has remained extremely limited, and only a few examples can be found of cases on discrimination by algorithms that have been decided or are currently pending before equality bodies or other semi-judicial bodies.

To the extent that cases on algorithms are brought before the national courts, they usually relate to other aspects of algorithmic decision making, such as accuracy, reliability or data protection. From the limited number of decisions available, it can further be deduced that some courts and equality bodies have found it problematic if decisions are automatically based on the outcome of the algorithm, without there being a 'human in the loop' to check and, if need be, correct the decision. Problems related to secrecy, transparency, explainability and information on the workings of an algorithm may also play a role in judicial proceedings. On matters of algorithmic stereotyping, biased data, the burden of proof in relation to algorithmic discrimination, or justification, no case law has been reported to exist.

Indeed, experts have reported that using judicial proceedings to enforce equality and non-discrimination principles in relation to algorithms is riddled with difficulties. In some countries, problems are experienced with non-discrimination law more generally, such as a lack of knowledge and expertise among courts and lawyers. This may discourage individuals from lodging non-discrimination cases, even more so in relatively abstract cases on discrimination by algorithms. However, even in countries where it is more common to bring judicial proceedings on matters of discrimination, this is considered much more complicated because of the particular characteristics of algorithms. As a result of the opacity and complexity of algorithms, it may be hard for individuals to know whether their use results in a form of unequal treatment and discrimination that is directly related to gender or other protected grounds. Relying on the concept of indirect discrimination might be a solution, but this concept is neither always sufficiently recognised on the national level nor sufficiently applied in national courts, in particular with regard to the shift of the burden of proof. Moreover, even if it is, experts note that it may be difficult to establish a *prima facie* case of indirect discrimination due to the particular evidentiary practices in any given country. In addition, the obscurity of algorithmic decision-making processes can make it difficult to obtain the necessary information to show a significant disadvantage for a protected group. As several experts have pointed out, this also shows that current equality and non-discrimination legislation focuses too strongly on offering legal protection in individual cases. This approach obliges individuals to show a case of indirect discrimination against a well-defined group, while many instances of algorithmic discrimination derive from more structural problems and could be better addressed at a collective level.

In summary, current national equality and non-discrimination laws are not regarded as being particularly well suited to dealing with the particular discrimination challenges of algorithmic decision making, and

courts and equality bodies have had little opportunity to provide further clarification and to fill gaps. In addition, although increasing public and scholarly attention is being paid to the ethical and legal aspects of algorithms, relatively few of these discussions focus on matters of equality and non-discrimination.

4 Enforcing algorithmic equality: Solutions and opportunities for gender equality and non-discrimination

4.1 Introduction

Overall, a rather bleak picture has emerged from this report so far: algorithms bring many challenges and difficulties in relation to discrimination, yet EU equality and non-discrimination laws are not particularly well equipped to deal with these, and on the national level, legal and judicial responses are usually tardy and limited. In addition, uncertainty remains as regards the combined ability of legal frameworks on equality, data protection and liability in place at EU and national level to adequately remedy algorithmic discrimination. Nevertheless, it is important to emphasise that, even if there is much reason for concern, there are still opportunities and positive dimensions to be mentioned, and there is considerable room for (legal) innovation. For that reason, this final chapter is forward-looking and addresses promising avenues for change and improvement.

First, to counterbalance the challenges, problems and risks that have been central to Chapters 1-3, section 4.2 presents some benefits of and opportunities of algorithms from an equality perspective. It is shown that even if algorithmic decision making raises many difficulties in relation to equality rights, algorithmic techniques can also be put to the service of the protection of gender equality and non-discrimination. In section 4.3, this chapter turns to describing a number of good practices that have been identified in the various European countries and that can offer inspiration when it comes to addressing and monitoring algorithmic discrimination. Finally, this chapter offers a mapping of possible solutions and tools to tackle algorithmic discrimination relying on a three-dimensional classification (law-knowledge-technology) (section 4.4) and proposes an integrated framework under the acronym 'PROTECT' (section 4.5).

4.2 Benefits and opportunities of algorithmic decision making

The problems, risks and challenges highlighted in this report with regard to algorithmic discrimination are significant, and it may not be easy to tackle them. Importantly, however, their very existence should not be regarded as a reason to ban the use of algorithms for decision making altogether – assuming that it would still be possible in light of the omnipresence of algorithms in modern societies. It is essential to keep in mind that there are opportunities to reduce and mitigate the risks and problems and to meet the challenges. In addition, there are many benefits and opportunities that are equally inherent to the use of algorithms as are the risks and problems, and that should not be ignored when striving to eradicate algorithmic discrimination. Although misusing technologies can contribute to reinforcing inequalities, relying on 'intelligent' machines, can also help in minimising or eliminating discrimination.⁵⁶⁹

First, algorithmic decision making can have significant advantages in terms of rationality, replicability, explainability, and transparency compared to the human brain. Indeed, it is important to note that the 'black box' metaphor that is often used to describe algorithms has a dual meaning.⁵⁷⁰ On the one hand, it 'can mean a system whose workings are mysterious; we can observe its inputs and outputs, but we cannot tell how one becomes the other', but on the other hand it 'can [also] refer to a recording device, like the data-monitoring systems in planes, trains, and cars'.⁵⁷¹ The 'opportunity' literature stresses how the first dimension applies to human brains at least as much as to algorithms: if the inner workings of algorithms are obscure, so, too, is the human decision-making process. In other words, the human brain, too, is a black box. Although it is tempting to assume that human reasoning is superior to machine thinking, this assumption is often simply untrue. If rule-based algorithms are correctly and reliably developed and

569 See eg Kleinberg, J and others (2019), 'Discrimination in the Age of Algorithms', available at: <https://arxiv.org/abs/1902.03731>.

570 See section 1.4.4 and see in particular Pasquale, F (2015), *The Black Box Society: The Hidden Algorithms Behind Money and Information* (Harvard University Press).

571 Pasquale, F (2015), *The Black Box Society: The Hidden Algorithms Behind Money and Information* (Harvard University Press 201) 3.

used, for example, they will produce vast numbers of decisions with a much higher degree of accuracy, consistency, precision and speed than would be possible if humans had to make such decisions.⁵⁷²

In addition to this, as has also been emphasised in section 1.4.1, human reasoning and decision making may be tainted by prejudice, biases or stereotypes.⁵⁷³ In traditional decision-making processes, such flaws and biases may be very difficult to detect and they can easily be masked.⁵⁷⁴ For example, it may well be true that a recruitment officer makes decisions on the suitability of candidates based on stereotyped and gendered views, either unconsciously or perhaps even intentionally.⁵⁷⁵ Although discriminatory decisions may easily result from this, it will be extremely difficult to identify them and prove that discrimination has been involved.⁵⁷⁶ In this respect, assistance by algorithms may present an important opportunity, especially because of the second side of algorithms as ‘black boxes’, which is their similarity to the black box of an aircraft. Where decision-making processes by human brains cannot be recorded, replicated or audited, such opportunities arise in relation to algorithmic decision-making process. The new possibilities might offer ways to better visualise, measure, detect and ultimately correct discriminatory biases if proper legal regulation and public policy is put in place. To continue with the recruitment example, it may be possible to prevent discrimination if a recruitment algorithm is used to support an HR officer’s decision-making process.⁵⁷⁷ The data used to train the algorithm can be carefully selected, cleaned and prepared to remove or reduce the risk of bias, and precise instructions may be given to the recruitment officers as to how to interpret the output offered by the algorithm and how to overrule it if necessary.⁵⁷⁸ Furthermore, if it turns out that, regardless of all such precautions, the algorithm produces discriminatory output, technical auditing tools may be used to pinpoint the exact point of failure of the algorithm and correct for it.⁵⁷⁹ Algorithms can also be developed and used to detect discrimination, both in the application of other algorithms,⁵⁸⁰ and in relation to discovering patterns of structural exclusion and disadvantage in specific sectors in society. Hence, there are many ways in which algorithmic decision making can be used to help reduce, rather than increase, the risk of discrimination.⁵⁸¹

These benefits and opportunities are also recognised by the national experts, who have noted several interesting examples of how algorithms could be used to this end. For example, for **Austria**, it has been explained in section 3.1.2.1 that the Austrian Labour Market Service (AMS) is developing an algorithm that, based on previous statistical labour market data, can be used to determine the future labour market chances of applicants.⁵⁸² Although various risks and problems of discrimination have been highlighted in

-
- 572 See e.g. Fagan, F and Levmore, S (2019) ‘The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion’ 93 *Southern California Law Review* 11; Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek* (Legal aspects of algorithms that make decisions. An exploratory study) (WODC/Utrecht University Utrecht) section 8.1.2.
- 573 See e.g. Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making: Opportunities and challenges* (Panel for the Future of Science and Technology (STOA) of the European Parliament) 10.
- 574 Compare Kleinberg, J and others (2019), ‘Discrimination in the Age of Algorithms’, available at: <https://arxiv.org/abs/1902.03731>, 2.
- 575 See further e.g. Raso, FA and others (2018), ‘Artificial intelligence & human rights: opportunities & risks’, The Berkman Klein Center for Internet & Society Research Publication Series: <https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights>, 43.
- 576 Compare Kleinberg, J and others (2019), ‘Discrimination in the Age of Algorithms’, 2.
- 577 Kleinberg, J and others (2019), ‘Discrimination in the Age of Algorithms’. See also Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making: Opportunities and challenges*, 10.
- 578 Compare e.g. Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’ 55 *Common Market Law Review* 1143, 1177; Corbett-Davies, S, Goel, S, and González-Bailón, S (2017), ‘Even imperfect algorithms can improve the criminal justice system. A way to combat the capricious and biased nature of human decisions’, *New York Times* (20 December 2017), <https://nyti.ms/2Deshvp>. On overruling and human-machine decision-making teams, see e.g. Fagan, F and Levmore, S (2019), ‘The impact of artificial intelligence on rules, standards, and judicial discretion’ 93 *Southern California Law Review* 11.
- 579 See e.g. Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’, 1177.
- 580 See e.g. Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’, 1175-6.
- 581 For additional algorithmic tools that may help to make recruitment processes more fair, see Raso, FA and others (2018), ‘Artificial intelligence & human rights: opportunities & risks’, The Berkman Klein Center for Internet & Society Research Publication Series: <https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights>, 44.
- 582 The launch was originally planned for mid-2020, but has been postponed to 2021.

relation to this system, the national expert has also remarked that the algorithm might be developed into a tool to collect valuable evidence on discriminatory structures within the Austrian labour market. Based on such information, it might be possible to develop data-based and targeted countermeasures.

Furthermore, algorithms may be consciously developed and used with the aim of tracing patterns of discrimination. In **Belgium**, for example, Article 42(1) of the Social Criminal Code allows labour inspectors who receive a complaint or an alert to use data mining and data matching systems to search for proof of discriminatory practice in the field of labour relations. If such a search yields *prima facie* evidence of discrimination, the *auditeur du travail/arbeidsauditeur* (specialised public prosecutor) can authorise the labour inspectorate to make mystery calls. Although some practical implementation difficulties have been reported⁵⁸³ and little is known about the workings of this system,⁵⁸⁴ the national expert considers that this use of big data analysis could be regarded as a way to support the fight against discrimination.⁵⁸⁵

In a similar vein, algorithms can track down discrimination in job advertisements, as can be seen in a pilot project that has been run in the **Netherlands**.⁵⁸⁶ Being trained on a quantity of texts containing age-based discrimination, an algorithm was asked to process approximately 2 million job advertisements to check whether they showed similar instances of discrimination. The algorithm found more than 40 000 textual elements that could be qualified as discriminatory. The results of this pilot could be used by the Labour Inspectorate to track down more forms of labour market discrimination, and as a basis for developing policies and legislation to counter this type of discrimination. In addition, such an analysis enables employers to remove the discriminatory passages from their job advertisements.⁵⁸⁷

Finally, it has been pointed out on the national level that the use of algorithms can sometimes help disadvantaged groups to obtain better access to certain benefits or social goods, including labour and education.⁵⁸⁸ In **Norway**, for example, the IT company Evry⁵⁸⁹ increased the recruitment of female employees after deploying a new AI-based recruitment process. The company stated that the AI used in the recruitment process guaranteed that decisions relied on more objective criteria than during regular human-led recruitment processes.⁵⁹⁰ In **Belgium**, the Flemish Government set up a centralised system to assign pupils to elementary schools in areas where educational institutions face a capacity problem.⁵⁹¹ The algorithm used was *inter alia* tasked to combat social segregation by establishing a so-called 'double quota', assigning places in schools to two groups, 'privileged' and 'disadvantaged' pupils, with the aim of increasing social diversity in schools.⁵⁹² This system implied that each school had a quota of places reserved to 'disadvantaged' pupils. This example shows how algorithms can be used to increase societal

583 In an interview conducted by Belgian expert Nathalie Wuiame with D Henrard, Labour Inspectorate, on 28 May, difficulties were mentioned in accessing the public database as well as the absence of some discriminatory categories such as maternity or transgender.

584 Although a circular has been adopted regarding the implementation of Article 42/1, its content is confidential, see Collège des procureurs généraux / College van procureurs-generaal (College of Public Prosecutors), *Circulaire/Omzendbrieven* COL. 15/2018 adopted on 20 December 2018 (confidential) available at: www.om-mp.be/fr/savoir-plus/circulaires.

585 The national expert also suggests that the Labour Inspectorate could, in the future, obtain more leeway in using big data, for example to identify signals of potential discrimination at sectorial level, without first having to rely on a specific alert or a complaint from a victim.

586 NOS (2019), 'Algoritmes moeten discriminatie in vacatureteksten opsporen' (Algorithms have to track down discrimination in job advertisements), 10 August 2019, available at: <https://nos.nl/artikel/2296998-algoritmes-moeten-discriminatie-in-vacatureteksten-opsporen.html>.

587 See also Raso, FA and others (2018), 'artificial intelligence & human rights: opportunities & risks', 44.

588 See also Castelluccia, C and Le Métayer, C (2019), *Understanding algorithmic decision-making: Opportunities and challenges* (Panel for the Future of Science and Technology (STOA) of the European Parliament) 10.

589 See Evry's website: <https://www.evry.com/no/>.

590 Gulbrandsen, C (2017), 'Rekrutterer ved hjelp av robot – ansetter flere kvinner' (Recruits using a robot – hires more women), *Dagens næringsliv* (7 December 2017) available at: www.dn.no/arbeidsliv/robotteknologi/utdannelse/evry-evry/rekrutterer-ved-hjelp-av-robot-ansetter-flere-kvinner/2-1-227334.

591 See Vervloesem, K (2020), 'In Flanders, an algorithm attempts to make school choice fairer' in Chiusi, F and others (2020) *Automating Society Report 2020* (Algorithm Watch and Bertelsmann Stiftung) 34, available at: <https://automatingsociety.org/wp-content/uploads/2020/10/Automating-Society-Report-2020.pdf>.

592 Disadvantaged children were defined according to two non-cumulative criteria: either their mother did not complete secondary education or the student was in receipt of a scholarship. See Vervloesem, K (2020), 'In Flanders, an algorithm attempts to make school choice fairer' in Chiusi, F and others (2020) *Automating Society Report 2020*.

equality in the distribution of valuable social goods such as education. However, the legal basis for this algorithmic system was revoked in 2019 and discussions on an alternative system have been on-going.⁵⁹³

Hence, even if only few concrete examples of such beneficial use of algorithms to tackle issues of discrimination have been provided, these clearly show that there is considerable potential.

4.3 Tackling algorithmic discrimination: a review of national good practice in European countries

Beyond the opportunities linked to the use of AI described above, European countries have also adopted good practices to address the risks and challenges arising from the use of algorithms. Although there is little specific legislation and case law dealing with algorithmic inequality and discrimination, as shown in Chapter 3, there is a flurry of activity in the field of policymaking, devising useful soft-law instruments such as ethical codes, and stimulating self-regulation in sectors where algorithms are actively used. With a few exceptions, most countries have either already adopted a national AI strategy or a white paper dealing with issues of AI and algorithms, or are in the process of doing so.⁵⁹⁴ Nearly all of these documents mirror the EU's efforts to increase the ethical, responsible and human-centric use of AI. This section maps out existing good practices in relation to algorithmic discrimination in European countries.

4.3.1 Monitoring algorithmic discrimination: examples of good practices and opportunities

Several monitoring practices can be identified in European countries, either by public bodies, civil society or by private organisations willing to ensure the ethical nature of their use of algorithms.

In **Belgium**, the VDAB (the public employment service in Flanders) has set up an ethical committee in charge of monitoring its algorithmic systems and ensuring that they are fair, ethical and do not lead to discrimination.⁵⁹⁵

In **Czechia**, the newly founded AI Observatory and Forum is supposed to serve as the 'Czech Republic's expert platform and forum for monitoring legal and ethical rules for artificial intelligence' and could thus play an important role in monitoring algorithmic discrimination.⁵⁹⁶

In **Finland**, the Non-Discrimination Ombudsman could play an important supervisory role. For instance, it may ask for information on allegedly discriminatory algorithms used by private companies – provided discrimination is not related to working life discrimination, which comes under the mandate of occupational safety officials.⁵⁹⁷ It also cooperates with the Data Protection Ombudsman to improve the use of the impact assessments by data users. For example, it asked the Financial Supervisory Authority to request changes in a private company's discriminatory algorithms after a decision by the Non-Discrimination and Equality Board that found a credit company guilty of algorithmic discrimination.⁵⁹⁸ Another Finnish body, the Council of Ethics, also monitors profiling techniques used in marketing/advertising, although the national expert has reported that a recent decision concerning Facebook ads shows some misunderstanding of the legal problems involved. A person had received advertisements on preventive HIV medication on his/her Facebook timeline, and reasoned s/he should not, on the basis of his/her social media use, be a

593 See Vervloesem, K (2020), 'In Flanders, an algorithm attempts to make school choice fairer' in Chiusi, F and others (2020) *Automating Society Report 2020*.

594 Countries where no such activities have been reported are **Austria, Belgium, Cyprus, Hungary, Iceland, Liechtenstein, Portugal, Romania, and Slovenia**.

595 See webinar organised by the Knowledge Centre and Data Society on AI in recruitment, available in Dutch at: <https://data-en-maatschappij.ai/en/news/data-date-2-ai-en-rekrutering>.

596 Available at: <http://observatory.ilaw.cas.cz>.

597 Interview conducted by Finnish expert Kevät Nousiainen with superinspector Tiina Valonen from the Non-Discrimination Ombud's office (27 May 2020).

598 See decision 216/2017 of 21 March 2018 of the Non-Discrimination and Equality Board, available at: www.yvtltk.fi/material/attachments/ytltk/tapausselosteet/2SVkNzOWF/YVTltk-tapausseloste-21.3.2018-luotto-moniperusteinen_syrjinta-S_L.pdf.

target for such advertising. The customer also argued that he/she had been profiled in a non-appropriate manner based on sexual orientation, but the Council of Ethics did not find the profiling discriminatory. The company using Facebook as a marketing platform commented that it had only used the open access Facebook tools for targeting the advertisement, i.e. the ‘interest in the subject’ of the Facebook user as opposed to information on sexual orientation. The argument was that preventive HIV medication could be useful for any sexually active person, irrespective of sexual orientation, and the advertisements were informative, neutral and non-discriminatory by nature. Facebook was also invited to comment but did not do so. Basing its decision on the marketing rules of the International Chamber of Commerce (ICC), the Council of Ethics decided that targeting based on a person’s use of Facebook (clicking, liking, sharing or demographic information) was not against ICC marketing rules, as sexual orientation had not been used for targeting.⁵⁹⁹

In **Poland**, besides the monitoring role assumed by civil society organisations such as the Panoptykon Foundation and the ePaństwo Foundation, the national draft policy on AI proposes the creation of two monitoring institutions – the Observatory of International AI Policy and Digital Transformation and the AI Observatory for the Labour Market – which, although their mandate is of general nature, could play a role in monitoring algorithmic discrimination.⁶⁰⁰

In **Spain**, the organisation Algorithm Watch assessed two algorithms: one called VioGén, used by the Spanish police to predict gender violence⁶⁰¹ and one called BOSCO, which was used by the Spanish Ministry for the Green Energy Transition, which had led to the rejection of over half a million applications for electricity subsidies.⁶⁰² In the first case, it criticised the algorithm’s critical failures and high rates of false negatives as well as the rubber-stamping attitude of the police in relation to algorithmic predictive output. In the second case, an association called Civio started legal proceedings because of the refusal of the Government to make the source code available so that the adequacy of the rejection decisions could be assessed.⁶⁰³ These monitoring activities show the important role that civil society organisations can play in the enforcement of equality law in cases of algorithmic discrimination by shining a light on the social impact of algorithmic decision-making. In the same vein, the Barcelona NGO Eticas Foundation⁶⁰⁴ has set up an Observatory of Algorithms with Social Impact (OASI),⁶⁰⁵ which gathers examples of high-impact algorithms around the world into a repository. This database could be a valuable source for further tests in relation to discrimination.

In **Sweden**, the Equality Ombudsman plays an important monitoring role, which has already led to two decisions so far, as discussed in section 3.4.2.1: one on bank loans where the automated calculation of future income was discriminatory on grounds of age⁶⁰⁶ and one on age discrimination in bank services using age identification.⁶⁰⁷

4.3.2 Addressing algorithmic discrimination: examples of good practices and opportunities

4.3.2.1 Recommendations and guidelines

Both public and private institutions have contributed to issuing guidelines and recommendations on the non-discriminatory use of AI. In **Denmark**, for example, the Danish Expert Group on Data Ethics

599 See <https://kauppakamari.fi/statement-archive/men-35-2019-mainoksen-kohdentaminen-sosiaalisessa-mediassa/>. The expert looked at cases decided in 2019 and 2020.

600 Ministry of Digital Affairs (2019), *Policy for the Development of Artificial Intelligence in Poland for 2019-2027*, 52.

601 See <https://algorithmwatch.org/en/story/viogen-algorithm-gender-violence/>.

602 See <https://algorithmwatch.org/en/story/spain-legal-fight-over-an-algorithms-code/>.

603 See <https://civio.es/tu-derecho-a-saber/2019/05/16/la-aplicacion-del-bono-social-del-gobierno-niega-la-ayuda-a-personas-que-tienen-derecho-a-ella/>.

604 See <https://eticasfoundation.org/about/>.

605 See <https://eticasfoundation.org/algorithms/>.

606 Equality Ombudsman, decision GRA 2017/80.

607 Equality Ombudsman, decision TIL 2018/22.

has highlighted four principles that are relevant to the fight against algorithmic discrimination: (1) self-determination, that is users' control over what their data is used for and in which contexts, (2) dignity, which must be respected in all instances of data processing, (3) equality and fairness, that is achieving a fair balance in data processing, the active prevention of harmful bias in data, the avoidance of discriminatory profiling, and the openness and revisability of de-biasing strategies and methods put in place, and finally (4) diversity, that is creating demographic diversity in relevant professional communities in order to guarantee that the needs, values and interests of various population groups are represented in the design of algorithmic systems.⁶⁰⁸ In 2019, the Danish Financial Supervisory Authority issued another set of recommendations, which asks companies using machine-learning techniques to actively reflect on the use of variables in algorithmic models, assess their outputs by involving experts with domain knowledge and consider de-biasing strategies if necessary.⁶⁰⁹

In **Estonia**, the AI task force has established diversity, non-discrimination and justice as one of the seven basic requirements for credible AI.⁶¹⁰ It supplements these guidelines with a checklist to help users assess compliance with these principles.

In **France**, the Defender of Rights and the National Commission on Informatics and Liberty (CNIL) have issued a set of recommendations to prevent algorithmic discrimination: (1) training and raising awareness of non-discrimination law and the discriminatory impact of algorithms among IT and data science professionals; (2) publicly supporting research into tools to evaluate biases in existing algorithms as well as interdisciplinary AI research; and (3) reinforcing legal obligations to inform citizens and consumers of the risks of discrimination in a clear and simple manner and developing methods to increase transparency and explainability of algorithms in the public and private sectors to the benefit of users but also third parties (unions, NGOs, civil society).⁶¹¹ The French think tank, Institut Montaigne, has made further recommendations spanning prevention, monitoring and assessment of algorithmic discrimination. In particular, it rejects the need to adopt a new law on algorithmic discrimination as well as state control over algorithms.⁶¹² It also recommends the dissemination of good practices to prevent the diffusion of algorithmic biases (internal codes of conduct, diversity in teams, etc.) as well as the involvement of management teams in companies when an algorithmic decision-making procedure is introduced or modified. Such management teams could play a role, for example, when a new variable is introduced, when algorithmic performance and fairness need to be balanced out, or when the definition of a standard of fairness or equality needs to be evaluated. The think tank also recommends training IT engineers and technicians as well as informing citizens about the risks of algorithmic discrimination. Its recommendations further concern the monitoring of algorithms through a regular testing of algorithms in use for potential discrimination (by analogy with clinical studies for drugs). In addition, the Institut Montaigne advises limiting the 'equity through ignorance' approach that consists in concealing sensitive variables from algorithms. While some banking and insurance companies suppress prohibited grounds from available variables used as actuarial factors in their algorithms, deep-learning algorithms can deduce protected grounds (for example gender) from other datapoints. These sensitive datapoints need to be visible to measure bias and discrimination. In this regard, the Institut Montaigne recommends comparing rates of false positives, that is checking that classification error rates are similar for each demographic subgroup so that no disadvantage arises from an algorithm that would be efficient and accurate for one group but

608 See <https://dataetiskraad.dk/sites/default/files/2020-02/Recommendations%20from%20the%20Danish%20Expert%20Group%20on%20Data%20Ethics.pdf>.

609 See https://www.dfsa.dk/Supervision/Fintech/Machine_learning_recommendations.

610 Kerikmäe, T (2019), *Autonoomsed intelligentsed tehnoloogiad ja õigusruum* (Autonomous Intelligent Technologies and Legal Framework), available at: https://docs.wixstatic.com/ugd/89dea2_c9baeef025a945c2a9585179074fd189.pdf.

611 Defender of Rights and CNIL (2020), *Algorithmes: prévenir l'automatisation des discriminations* (Algorithms: preventing automated discrimination) joint report, 7 available at: www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-en-num-16.07.20.pdf and ENA (2019), *Rapport collectif sur commande de la mission Etalab, Ethique et responsabilité des algorithmes publics* (Collective report commissioned by the Etalab mission on the Ethics and Responsibility of Public Algorithms) (ENA Year 2018-2019, Molière, June 2019) available at: www.etalab.gouv.fr/wp-content/uploads/2020/01/Rapport-ENA-Ethique-et-responsabilite-C3%A9-des-algorithmes-publics.pdf.

612 Institut Montaigne (2020), *Algorithmes: contrôle des biais s.v.p.* (Algorithms: Bias control) (March 2020) available at: www.institutmontaigne.org/ressources/pdfs/publications/algorithmes-controle-des-biais-svp.pdf.

not the other. The organisation also suggests the creation of public test databases that companies could use to assess the discriminatory risks arising from their algorithms. Finally, it proposes that a higher level of scrutiny is applied to high-impact algorithms, recommends the creation of labels and certifications to foster trust when it comes to algorithmic decision making and supports the development of auditing capacities for high-impact algorithms.

The **Italian** white paper on 'AI at the service of the citizen', published in 2018 by the government body, Agency for Digital Italy, contains recommendations for the public sector on how to best exploit the opportunities offered by AI and limit its negative impact.⁶¹³ In relation to algorithmic discrimination, it recognises several conditions for responsible use of AI, including the importance of good and unbiased data, the responsibility and accountability of those who rely on algorithms for decision making, and the need to reconcile several principles such as the transparency of administrative acts and algorithmic procedures, the protection of personal data, the protection of copyrights and the right to privacy. In response, the Italian white paper recommends (1) the creation of a national platform dedicated to carrying out tests to assess the potential discriminatory impact of algorithms before their use by the public administrations, and (2) publicising the parameters of algorithms used by public administrations so as to allow reproducibility, evaluation and verifiability (with citizens' privacy and security as limiting principles).

The **Norwegian** Board of Technology has addressed a number of broad generic recommendations to the Government, which concern the issue of algorithmic discrimination but also other AI-related challenges.⁶¹⁴ The recommendations start from principles that have already been mentioned above, such as open public data, support for research into AI, the right to an explanation, auditing algorithms, open algorithms in the public sector, a 'digital social contract' that would give citizens real opportunities to control if and how their data is shared, and an 'ethics by design approach' specifically in relation to algorithmic bias.

In **Poland**, several foundations, such as the Panopticon Foundation, the ePaństwo Foundation and the Foundation Centrum Cyfrowe, have issued recommendations on the use of algorithmic decision making in the public sector.⁶¹⁵ These relate to legal, procedural and practical safeguards to ensure full transparency, evidence-based impact assessments of human rights-related risks, expected social benefits, adequacy and proportionality, means to prevent rubber-stamping of algorithmic output in public administrations, and additional legal safeguards where algorithmic decision making involves the processing of sensitive personal data or information related to discrimination grounds (judicial or administrative redress procedures, right to compensation, reversal of burden of proof).

The **Swedish** Disability Rights Federation has issued various proposals for inclusive AI. For example, it has called for a more active inclusion of persons with disabilities in relation to both public procurement and AI research and development as well as awareness raising and knowledge development concerning AI and discrimination.⁶¹⁶

613 See AGID (Agenzia per L'Italia Digitale) (2018), *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino* (White paper on AI at the service of the citizen) (Version 1.0) available at: <https://ia.italia.it/assets/librobianco.pdf>.

614 See NBT (2018), *Kunstig Intelligens, muligheter, utfordringer og en plan for Norge* (AI, possibilities, challenges and a plan for Norway) (September 2018) available at: <https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf>, 52-3 and 74.

615 Niklas, J, Sztandar-Sztanderska, K and Szymielewicz, K (2015), *Profiling the unemployed in Poland: social and political implications of algorithmic decision making* (Eliza Jakubiak tr, report by Fundacja Panoptikon, Warsaw) available at: https://panoptikon.org/sites/default/files/leadimage-biblioteka/panoptikon_profiling_report_final.pdf. See also <https://epf.org.pl/pl/2019/05/13/maszyny-sluzbie-panstwa-algorytmy-wplywaja-zycie-obywateli-obywatelek-polsce-wybranych-krajach-regionu/> and <https://epf.org.pl/pl/2019/10/28/manifest-epanstwa-postulaty-fundacji-epanstwo-na-kadencje-sejmu-2019-2023/>.

616 Swedish Disability Rights Federation (2019), *Förslag för inkluderande AI* (Proposals for inclusive AI), available at: <https://funktionsratt.se/funktionsratt-ratten-att-fungera-i-samhallet-pa-lika-villkor/ett-samhalle-for-alla/1148-2/artificiell-intelligens/forslag-for-inkluderande-ai/>.

Finally, in the **United Kingdom**, the data ethics group of the Alan Turing Institute⁶¹⁷ has published detailed guidance on ethics and artificial intelligence.⁶¹⁸ In essence, it sets out three sets of values or principles that should govern the planning, design and implementation of any project involving AI and considers how these may be implemented. These include: determining the ethical scope of AI projects; a framework to ensure that projects are 'bias-mitigating, non-discriminatory and fair' and to 'safeguard public trust'; and a framework to set up transparent processes of design and implementation.

In summary, the policy objectives that are mentioned most often by think tanks, NGOs and similar institutions and organisations concern the ethical, reliable, non-discriminatory and human-centric use of algorithms, as well as the transparency and explainability of algorithms and algorithm-based decision making. Recommendations to invest in training and awareness raising about the potential risks of algorithmic discrimination are also sometimes made, as can be seen in **France**,⁶¹⁹ the **Netherlands**,⁶²⁰ **Slovakia**⁶²¹ and the **United Kingdom**.⁶²² Other possibilities that are regularly suggested are introducing AI impact assessments (**Germany**,⁶²³ the **Netherlands**,⁶²⁴ **Poland**,⁶²⁵ **Spain**⁶²⁶), offering protection against distortion, manipulation and other misuse (**Germany**),⁶²⁷ looking for man-machine cooperation in industrial production and AI systems (**Germany**),⁶²⁸ testing, certification and/or authorisation procedures (**France**,⁶²⁹ **Germany**,⁶³⁰ the **Netherlands**⁶³¹), and different systems for auditing (**France**,⁶³²

617 <https://www.turing.ac.uk/research/interest-groups/data-ethics-group>.

618 Leslie, D (2019), *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector* (The Alan Turing Institute) available at: www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf.

619 Defender of Rights and CNIL (2020), *Algorithmes: prévenir l'automatisation des discriminations* (Algorithms: preventing automated discrimination) joint report 7 available at: www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-en-num-16.07.20.pdf.

620 The Netherlands Government's policy brief on offering guarantees against the risks of data-analyses by public bodies of October 2019, available at: www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2019Z19084&did=2019D39751, contains an annex setting out *Guidelines for the application of algorithms by public bodies*. These guidelines emphasise the need for awareness of the risks of using algorithms, including a risk of discrimination and stigmatisation, in relation to all development and use of algorithms by the government.

621 *Action plan for the digital transformation of Slovakia for 2019–2022*, adopted by a Government Resolution No 337/2019 of 3 July 2019, available at: www.vicepremier.gov.sk/wp-content/uploads/2019/10/AP-DT-English-Version-FINAL.pdf.

622 The UK's Data Ethics Framework provides principles and guidance for using algorithms in the public sector and includes a recommendation to ensure awareness of the relevant legislation including the Equality Act 2010; see www.gov.uk/government/publications/data-ethics-framework.

623 Federal Government (2020), <https://www.ki-strategie-deutschland.de/home.html>.

624 Government of the Netherlands (2019), *Strategisch Actieplan Artificiële Intelligentie (SAPAI)* (Strategic Acton Plan Artificial Intelligence (SAPAI)) (October 2019) available at: www.rijksoverheid.nl/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie. This may be further developed as part of the Toolbox Ethics by Design; see <https://www.digitaleoverheid.nl/actielijn/ethics-by-design/>.

625 Ministry of Digital Affairs (2019), *Polityka Rozwoju Sztucznej. Inteligencji w Polsce na lata 2019 – 2027* (Policy for the Development of Artificial Intelligence in Poland for 2019 – 2027) 103.

626 Guidelines on the Adaptation of AI adopted by the Spanish Agency for Data Protection, *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial* (Adaptation to the GDPR of treatments that incorporate Artificial Intelligence) available at: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf.

627 Federal Government (2018), 'Strategie Künstliche Intelligenz der Bundesregierung' (National Strategy on Artificial Intelligence), available at: www.ki-strategie-deutschland.de/home.html; Federal Government (2020), 'Künstliche Intelligenz (KI) ist ein Schlüssel zur Welt von morgen' (Artificial Intelligence is a Key to Tomorrow's World), available at: <https://www.ki-strategie-deutschland.de/home.html>.

628 <https://testing-ai.gi.de/>.

629 E.g. Institut Montaigne (2020), *Algorithmes: contrôle des biais s.v.p.* (Algorithms: Bias check, please) (March 2020) available at: www.institutmontaigne.org/ressources/pdfs/publications/algorithmes-contrôle-des-biais-svp.pdf.

630 Data Ethics Commission (2019), *Expertise*, available at: www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenethik/kommission/arbeitssergebnisse-der-dek/arbeitssergebnisse-der-dek-node.html; Federal Government (2020), 'Künstliche Intelligenz (KI) ist ein Schlüssel zur Welt von morgen'; see also <https://testing-ai.gi.de/>.

631 Government of the Netherlands (2019), *Strategisch Actieplan Artificiële Intelligentie (SAPAI)* (Strategic Acton Plan Artificial Intelligence (SAPAI)) (October 2019).

632 See e.g. Institut Montaigne (2020), *Algorithmes : contrôle des biais s.v.p.* (Algorithms: Bias check, please) (March 2020).

Germany,⁶³³ **Italy**,⁶³⁴ the **Netherlands**,⁶³⁵ **Poland**,⁶³⁶ the **United Kingdom**⁶³⁷). In **Germany**, the creation of a 'diversity seal' or seal of approval for algorithmic systems that comply with established standards of data protection, equality and non-discrimination has been suggested.⁶³⁸

4.3.2.2 Voluntary codes of conduct, co-regulation and self-regulation

There is also considerable activity in developing soft law and self-regulation or co-regulation on matters related to gender equality and non-discrimination and algorithms. Instruments that are available or in the process of being developed often come in the form of guidelines, toolboxes, manuals or frameworks for the ethical development and use of algorithms. The development of ethical codes or codes of practice is also popular. Sometimes these instruments concern the use of AI by public authorities only, such as the toolbox for 'ethics by design' that is under development in the **Netherlands**,⁶³⁹ or the data ethics framework that has been devised in the **United Kingdom**.⁶⁴⁰ In other cases, they are specifically designed to be used in specified private sector. In the **Netherlands** and **Slovakia**, for instance, there are specific ethics codes and manuals for the ICT sector;⁶⁴¹ **Poland** has a code of good practice for processing data in the banking sector;⁶⁴² in **Luxembourg**, the Supervisory Commission of the Financial Sector devised an ethical code;⁶⁴³ the **Norwegian** trade unions, employers' organisations and large ICT companies have agreed on a declaration on the use of AI in the employment sector;⁶⁴⁴ and in **Czechia**, a platform for artificial intelligence has been established by the Czech industry and transport sectors.⁶⁴⁵

Some of these soft-law and self-regulatory instruments specifically aim to eradicate issues of algorithmic discrimination or may contribute to combating such discrimination. The **Polish** code of good practice for processing data in the banking sector, for example, limits the use of automated decision making and profiling,⁶⁴⁶ while the **Norwegian** declaration on the use of AI in labour and the **Slovakian** manual for IT companies on the deployment of AI explicitly mention the aim of non-discrimination.⁶⁴⁷ The **French** Defender of Rights has developed a toolkit for employers and HR services to prevent

633 Data Ethics Commission (2019), *Expertise*; Federal Government (2020), 'Künstliche Intelligenz (KI) ist ein Schlüssel zur Welt von morgen'; see also <https://testing-ai.gi.de/>.

634 AGID (Agenzia per L'Italia Digitale) (2018), *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino* (White paper on AI at the service of the citizen) (Version 1.0) available at: ia.italia.it/assets/librobianco.pdf.

635 Government of the Netherlands (2019), *Strategisch Actieplan Artificiële Intelligentie (SAPAI)*. See also the Netherlands Government's policy brief of 8 October 2019 on offering guarantees against the risks of data-analyses by public bodies, which contains an annex setting out *Guidelines for the application of algorithms by public bodies*; available at: https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2019Z19084&did=2019D39751.

636 Ministry of Digital Affairs (2019), *Polityka Rozwoju Sztucznej. Inteligencji w Polsce na lata 2019 – 2027* (Policy for the Development of Artificial Intelligence in Poland for 2019-2027) 103.

637 <https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf>.

638 State Anti-Discrimination Body of Berlin (2019), 'Algorithmen und das Recht auf digitale Gleichbehandlung. Dokumentation der Fachwerkstatt der Landesstelle für Gleichbehandlung – gegen Diskriminierung' (Algorithms and the right to digital equality. Documentation of the specialist workshop), available at: www.berlin.de/sen/lads/ueber-uns/materialien/.

639 Government of the Netherlands (2019), *Strategisch Actieplan Artificiële Intelligentie (SAPAI)*; see <https://www.digitaleoverheid.nl/actielijn/ethics-by-design/>.

640 <https://www.gov.uk/government/publications/data-ethics-framework>.

641 For the **Netherlands**, see NLdigital (2019), 'Nieuwe ethische gedragscode AI voor ICT-sector' (New ethics code AI for ICT Sector) (21 March 2019) available at: www.nldigital.nl/news/nieuwe-ethische-gedragscode-ai-voor-ict-sector/; an English language version of the code is available at: <https://www.nldigital.nl/wp-content/uploads/2019/04/NEDERLAND-ICT-ETHICAL-CODE-FOR-ARTIFICIAL-INTELLIGENCE.pdf>; for **Slovakia**, see: <https://www.vicepremier.gov.sk/wp-content/uploads/2020/03/Dielo2-Manual.pdf>.

642 <https://uodo.gov.pl/pl/file/2362>.

643 CSSF – Commission de Surveillance du Secteur Financier (2018), *White Paper on Artificial Intelligence. Opportunities, risks and recommendations* available at: www.cssf.lu/en/document/white-paper-artificial-intelligence-opportunities-risks-and-recommendations-for-the-financial-sector/.

644 Minister of Local Government and Modernisation (2019), 'Erklæring om ansvarlig bruk av kunstig intelligens i arbeidslivet' (Declaration on the responsible use of artificial intelligence in working life. A Declaration signed by various companies and organizations in private and public sector in Norway) available at: www.regjeringen.no/contentassets/0e36c85f9cfe143a5b626c53cf292cb3b/negotia_telenor_ansvarlig-bruk-av-kunstig-intelligens-i-arbeidslivet.pdf.

645 See www.spcr.cz/images/Z%C3%A1kladn%C3%AD_dokument_platforma_pro_AI_SP%C4%8CR.pdf.

646 See <https://uodo.gov.pl/pl/file/2362>.

647 **Norway**: Minister of Local Government and Modernisation (2019), 'Erklæring om ansvarlig bruk av kunstig intelligens i arbeidslivet'; **Slovakia**: <https://www.vicepremier.gov.sk/wp-content/uploads/2020/03/Dielo2-Manual.pdf>.

algorithmic discrimination during recruitment processes.⁶⁴⁸ Other such toolkits have been developed by private companies, for example IBM's open-source toolkit to detect and mitigate bias in datasets and algorithms.⁶⁴⁹ Moreover, the **Danish** Expert Group on Data Ethics has proposed that companies take a 'data ethics oath' by which they adhere to a responsible use of data and take active measures to ensure that their use of data is not discriminatory or biased on the grounds of sex, ethnicity or other protected social groups.⁶⁵⁰ The **Dutch** NVP, a network of HR professionals, developed a new recruitment code in March 2020, which states that where AI or algorithms are used in recruitment, the means used should be transparent and validated and the possible risks and shortcomings must be clear. The NVP recruitment code is not binding, but sets the standard for rules that companies and organisations should take into account when recruiting new personnel.⁶⁵¹

In self-regulation instruments and codes of conduct, emphasis is usually also placed on the need for human-centred AI and transparency (e.g. in **Malta**,⁶⁵² the **Netherlands**⁶⁵³ and **Norway**⁶⁵⁴). Furthermore, these instruments may help to guarantee adequate testing of algorithms (e.g. by means of the collaborative scientific platform TransAlgo in **France**)⁶⁵⁵ or may be vehicles to ensure certification (e.g. the **Maltese** AI certification programme).⁶⁵⁶

Such self-regulatory instruments are often developed or used by public-private coalitions or platforms. These may allow for exchanges of knowledge and best practice, and help to identify and prevent instances of algorithmic discrimination. In **Hungary**, for example, there is an Artificial Intelligence Coalition,⁶⁵⁷ in the **Netherlands**, the National Coalition AI,⁶⁵⁸ the Platform for the Information Society⁶⁵⁹ and the Alliance for Artificial Intelligence⁶⁶⁰ have been established; in **Norway**, there is a joint forum on AI,⁶⁶¹ and in **Slovakia**, there is a national platform for research and utilisation of AI (Slovak AI).⁶⁶²

4.3.2.3 Role of public bodies

In most countries, a variety of public bodies are involved in policymaking and the regulation of algorithms. These are often Government offices or ministries, but independent technology agencies and AI bodies may also play a role. Active examples of this include the Committee on Information Systems and Freedom (CNIL) in **France**,⁶⁶³ the digital ombudsman in **Italy**,⁶⁶⁴ the Digital Innovation Authority in **Malta**,⁶⁶⁵ and

648 Défenseur des Droits (2015), *Guide: Recruter avec des outils numériques sans discriminer* (Guide: Recruiting with digital tools without discrimination) available at: <https://www.defenseurdesdroits.fr/fr/outils/recruter-avec-des-outils-numeriques-sans-discriminer>.

649 Varshney, KR (2018), 'Introducing AI Fairness 360', *IBM Research Blog*, available at: www.ibm.com/blogs/research/2018/09/ai-fairness-360/.

650 Danish Expert Group on Data Ethics (2020), *Data for the Benefit of the People*, available at: <https://dataetiskraad.dk/sites/default/files/2020-02/Recommendations%20from%20the%20Danish%20Expert%20Group%20on%20Data%20Ethics.pdf>, Recommendation 2.

651 NVP (2020), *Sollicitatiecode* (Recruitment Code), available at: www.nvp-hrnetwerk.nl/sollicitatiecode/.

652 Malta (2018) *Ethical AI Framework*, available at: https://malta.ai/wp-content/uploads/2019/10/Malta_Towards_Ethical_and_Trustworthy_AI_vFINAL.pdf.

653 Government of the Netherlands (2019), *Strategisch Actieplan Artificiële Intelligentie (SAPAI)*.

654 Minister of Local Government and Modernisation (2019), 'Erklæring om ansvarlig bruk av kunstig intelligens i arbeidslivet'.

655 See www.transalgo.org/.

656 See <https://mdia.gov.mt/wp-content/uploads/2018/10/ITAS.pdf>.

657 Artificial Intelligence Coalition (*Mesterséges Intelligencia Koalíció*), established in October 2018, see <https://digitalisjoletprogram.hu/hu/tartalom/mesterseges-intelligencia-koalicio>.

658 The Netherlands AI Coalition (*Nederlandse AI Coalitie* – NL AIC), see www.vno-ncw.nl/standpunten/artificial-intelligence-ai.

659 Platform for the Information Society (ECP), see <https://ecp.nl>.

660 Alliance for Artificial Intelligence (ALLAI), see <https://allai.nl/>.

661 See www.difi.no/.

662 See <https://slovak.ai/2019/06/28/slovak-ai-presentation/?lang=en>.

663 See www.cnil.fr.

664 Based on the Digital Administration Code, see *Codice dell'Amministrazione Digitale, Decreto Legislativo 7 marzo 2005, n. 82* (Legislative Decree no. 82 of 7 March 2005), available at: <https://docs.italia.it/italia/piano-triennale-ict/codice-amministrazione-digitale-docs/it/v2017-12-13/index.html>.

665 In **Malta**, this authority is responsible for drafting and implementing the AI Strategy; for more information, see https://malta.ai/wp-content/uploads/2019/11/Malta_The_Ultimate_AI_Launchpad_vFinal.pdf.

the Centre for Data Ethics and Innovation in the **United Kingdom**.⁶⁶⁶ Data protection authorities, equality bodies, ombudspersons and human rights institutes may play an equally important role, either in an advisory capacity or/and in dealing with individual complaints brought about by experiences of algorithmic discrimination.⁶⁶⁷

Since the use of algorithms is often strongly determined by the needs of a specific sector, sectoral bodies (both public and private) can also be relevant to the process of enforcing relevant gender equality and non-discrimination rules where algorithmic applications are used. For example, in **Belgium** and **Cyprus**, national experts suggest that public agencies in the field of employment could be further involved in relation to algorithms and discrimination in the future.⁶⁶⁸ For **Iceland**,⁶⁶⁹ **Poland**⁶⁷⁰ and **Sweden**, national experts reported that social partners, trade unions and employers' organisations are playing a role or are expected to do so in the future.

In relation to gender inequality in the field of goods and services, the activities of consumer authorities have been noted in **Norway**,⁶⁷¹ **Poland**⁶⁷² and **Sweden**.⁶⁷³ There is also an important role for independent advisory councils, such as the Economic and Social Council in **Bulgaria**,⁶⁷⁴ the National Digital Council in **France**,⁶⁷⁵ the **Netherlands** Scientific Council for Government Policy,⁶⁷⁶ the **Norwegian** Board on Technology⁶⁷⁷ and the Artificial Intelligence Council in the **United Kingdom**.⁶⁷⁸

Finally, in many countries, there are dedicated think tanks, research centres and institutes that are involved in research to identify the effects of the use of AI and algorithms on society, as well as making proposals and recommendations to Governments to ensure ethical development and use of algorithms.

4.3.2.4 Public-private alliances and ethical platforms

Numerous national and transnational alliances have been established between private and public stakeholders to respond to ethical and legal challenges linked to the use of AI. For example, a platform for artificial intelligence was founded in 2018 by the **Czech** Confederation of Industry and Transport (the most important employers' union) in order to respond to the ethical, social, economic and legal implications of AI, monitor its impact on the labour market and participate in law-making at the national and European

666 See Centre for Data Ethics and Innovation at: www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation/about#:~:text=The%20Centre%20for%20Data%20Ethics,the%20benefits%20of%20these%20technologies.&text=We%20are%20a%20connector%20between,mandate%20to%20advise%20the%20government.

667 See further section 3.4.2.

668 In **Belgium**, for instance, the Labour Inspectorate can use data mining and data matching techniques to gather evidence of discriminatory practices in the labour market and the expert suggests that this competence could be widened in the future. See also section 4.2. In **Cyprus**, the Committee on Gender Equality in Employment and Vocational Training, an independent agency of the Ministry of Labour, can examine gender discrimination complaints in the context of employment/vocational training, offer free legal advice to victims and assist them in pursuing their claim at the courts. It can also organise information and awareness campaigns. The national expert suggests that it could play a useful role in addressing gender-based algorithmic discrimination at work.

669 ASI – the Icelandic Confederation of Labour, see <https://www.asi.is/english/>.

670 On the role to be played by the social partners, see e.g. Żebrowski, P (2019), *Kadrowe algorytmy – komputer może dyskryminować pracowników* (Personnel algorithms – computers can discriminate against employees) (Wyd. WKP, Sekcja, Kadry i BHP).

671 Consumer Authority (CA), see www.forbrukerradet.no/forside/om-oss/.

672 Office of Competition and Consumer Protection (UOKiK); see www.uokik.gov.pl/powiadomienia.php?news_id=13053;www.rpo.gov.pl/pl/content/tk-szczeg%C3%B3%C5%82y-profilowania-pomocy-dla-bezrobotnego-maj%C4%85-by%C4%87-uregulowane-w-ustawie;www.rpo.gov.pl/pl/content/liczenie-usmiechow-pracownikow-banku-pko-bp-budzi-watpliwosci-rpo.

673 Sandberg, A (ed) (2017), *Kunskapsöversikter inom det konsumentpolitiska området* (Knowledge reviews in the area of consumer policy), Konsumentverket (Swedish Consumer Agency) (2017:1) available at: <https://www.konsumentverket.se/globalassets/publikationer/produkter-och-tjanster/gemensamt/rapport-2017-1-antologi-konsumentverket.pdf>.

674 See Economic and Social Council, Икономически и социален съвет/ Ikonomiczeski I sozialen savet, via www.esc.bg/.

675 See <https://cnnumerique.fr/index.php/le-conseil>.

676 See <https://english.wrr.nl>.

677 See <https://teknologiradet.no/en/the-council-and-council-members/>; for its activities, see e.g. <https://teknologiradet.no/en/category/artificial-intelligence-ai-en/>.

678 See www.gov.uk/government/groups/ai-council.

level.⁶⁷⁹ In **Sweden**, the AI Sustainability Centre, a public-private partnership, was established in 2018 to create a world-leading multidisciplinary hub to address the scaling of AI in broader ethical and societal contexts.⁶⁸⁰ In the **United Kingdom**, Hackney Council and four other councils have partnered with the company Xanthura and, in consultation with the Information Commissioner's Office, to eliminate bias in algorithmic systems used to identify children in potential need of child services support. It has developed information governance processes including a series of ethical questions that were mapped to various stages of the implementation process.⁶⁸¹

In addition, in some countries, such as **France**,⁶⁸² **Italy**,⁶⁸³ the **Netherlands**⁶⁸⁴ and **Poland**,⁶⁸⁵ (ethical) platforms have been created (or are in the process of being established) for public and private developers and users of AI to help them exchange knowledge and good practice. Other countries, such as **Lithuania**⁶⁸⁶ and **Slovakia**⁶⁸⁷ favour an ethical committee that can help to evaluate and approve AI projects. Such ethical committees can also be established for specific sectors. For example, in **Belgium**, the Public Employment Service in Flanders has created an ethical committee for the development, experimentation and implementation of AI systems.⁶⁸⁸

4.3.2.5 Other examples of good practice

In **Italy**, the main trade union of the country, the CGIL, has campaigned to 'contract the algorithms', arguing that if algorithms impact the organisation of work, they must be included in the collective agreements and subject to bargaining in order to eliminate risks of discrimination.⁶⁸⁹

The **Dutch** normalisation institute (the Royal Netherlands Standardisation Institute, NEN), has set up a committee on artificial intelligence and big data to influence ISO standards in restricting bias, risk management and reliability of AI.⁶⁹⁰

Furthermore, various knowledge promotion, awareness-raising and funding initiatives have been established in recent years. The **Estonian** action plan on AI, for example, promotes expert knowledge and funds an awareness-raising and training programme for the wider public.⁶⁹¹ In the **UK**, the AI Law Hub provides information updates, commentaries and useful resources on legal issues, including discrimination, arising from AI.

679 Founding document available at: www.spcr.cz/images/Z%C3%A1kladn%C3%AD_dokument_platforma_pro_AI_SP%C4%8CR.pdf.

680 See www.aisustainability.org/about-us/.

681 A full set of responses can be found at: <https://old.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2015/inquiry9/publications/>.

682 See e.g. collaborative scientific platform Transalgo: <https://www.transalgo.org/>. It is also suggested that there could be specific platforms for different sectors; see Ministry on Innovation and Digital Issues and Ministry of Education (2017), *Rapport de synthèse, France Intelligence Artificielle* (Summary Report, France Artificial Intelligence) available at: www.cnil.fr/sites/default/files/atoms/files/rapport_de_synthese_france_ia.pdf, 4.

683 AGID (Agenzia per l'Italia Digitale) (2018), *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino* (White paper on AI at the service of the citizen) (Version 1.0) available at: ia.italia.it/assets/librobianco.pdf.

684 See e.g. Platform for the Information Society (ECP), see <https://ecp.nl/>; the Netherlands AI Coalition (*Nederlandse AI Coalitie* – NL AIC), see <https://www.vno-ncw.nl/standpunten/artificial-intelligence-ai>; Alliance for Artificial Intelligence (ALLAI), see <https://allai.nl/>.

685 See e.g. Forum Cyfrowego Rozwoju (Digital Development Forum); Internet Governance Forum, available at: <https://igf.nask.pl/>.

686 *Lithuanian Artificial Intelligence Strategy* (March 2019) available at: [http://eimin.lrv.lt/uploads/eimin/documents/files/DI_strategija_ENG\(1\).pdf](http://eimin.lrv.lt/uploads/eimin/documents/files/DI_strategija_ENG(1).pdf).

687 *Action plan for the digital transformation of Slovakia for 2019–2022*, adopted by a Government Resolution No. 337/2019 of 3 July 2019, available at: <https://www.vicepremier.gov.sk/wp-content/uploads/2019/10/AP-DT-English-Version-FINAL.pdf>.

688 Information based on interview by national expert Nathalie Wuiame with G Vanhumbeeck (Director for Innovation), K Scheerlinck (AI team leader) and V Buekenhout (Data Protection Manager), VDAB, online Teams meeting (12 June 2020).

689 See Baratta, L (2017), 'La proposta della Cgil: l'algoritmo di Deliveroo e Foodora crea discriminazioni, contrattiamolo' (Interview. The proposal of Cgil: the algorithm of Deliveroo and Foodora creates discrimination, let's contract it), *Linkiesta* (14 December 2017) available at: www.linkiesta.it/2017/12/la-proposta-della-cgil-lalgoritmo-di-deliveroo-e-foodora-crea-discriminazioni/.

690 NEN (2020), *Artificiële Intelligentie and Big Data*, available at: www.nen.nl/Normontwikkeling/Doe-mee/Normcommissies-en-nieuwe-trajecten/AI.htm.

691 EUR 300 000; planned to start in Spring 2020.

4.3.3 The diversity question in relevant professional and educational communities

4.3.3.1 Introduction

As was shown in section 3.2.6, in nearly all European countries, there is a lack of diversity in the workforce in STEM fields, as well as a persistent gender gap. This is considered problematic, since algorithms are based on statistical relationships and their outcomes reflect majority standards; hence, underrepresentation of certain groups bears the risk of excluding minority views and perspectives when developing and training algorithms.⁶⁹² It is for that reason that it is not only important for experts working with algorithms and data to be aware of these risks of bias, but also that educational and professional STEM communities themselves should be composed in a diverse and balanced manner.

Indeed, in nearly all countries studied in this report, public and private actors have taken steps towards increasing the diversity of educational and professional communities in STEM. While such steps can include certain types of quota, many other types of legally binding measures can be deployed to support gender equality in the digital sector, which would in turn reduce the risk of gender-based algorithmic discrimination by lessening risks that gender biases creep into algorithmic systems.⁶⁹³ Indeed, most countries have adopted a national gender equality strategy which takes into account the underrepresentation of women in technology education and professions, and aims to address the issue by awareness-raising, empowerment and training initiatives and programmes. Considerable efforts are dedicated to increasing women's participation in STEM education, as is discussed in section 4.4.3.2. Such measures are often aimed at familiarising girls and young women with existing job opportunities in the tech and IT sectors and at providing alternative role models to combat the enduring gender segregation in digital educational and professional sectors. In addition, as section 4.4.3.3 will show, there are many measures and programmes aimed at increasing the participation of women in STEM professions, which for instance offer opportunities to create support networks, or help to raise awareness among recruiters of the problem of underrepresentation of women. While some funding and dedicated policies can be identified, most of the measures discussed below are voluntary, non-binding or promotional. Legally binding policies such as quotas or other positive action measures, for instance special funding, are relatively scarce.

4.3.3.2 Efforts to increase gender diversity in STEM education

Many initiatives to advance female participation in STEM education focus on education. A considerable number of countries have relevant Government or trade union supported programmes and initiatives in place to promote gender diversity in tech education. Although such programmes are sometimes Government-supported, in many cases they are organised or supported by civil society or private companies, and they may be either national or local in nature. Strong positive action measures, such as introducing a quota for female participation in tech studies, have not been reported. To the contrary, in **Finland**, for example, the expert has observed that universities are cautious about the use of positive action in student recruitment policies due to the prohibition in the Finnish Act on Equality.⁶⁹⁴

692 Hatzopoulos, V (2020), 'Η διακυβέρνηση του κυβερνοχώρου' (The governance of cyberspace) in Stephanou, K (ed.), Το σύστημα διακυβέρνησης της Ευρωπαϊκής Ένωσης (The system of governance of the European Union) (Athens, Nomiki Viviothiki) 402.

693 On positive action in employment in Europe, see McCrudden, C (2019), *Gender-based positive action in employment in Europe: A comparative analysis of legal and policy approaches in the EU and EEA* (EELN, European Commission) available at: <https://www.equalitylaw.eu/downloads/5008-gender-based-positive-action-in-employment-in-europe-pdf-1-9-mb>.

694 The Equality Ombudsman has given several opinions on gender quotas in higher education and has found that such quotas violate Section 7 of the Act on Equality. The latest of these opinions, dated 7 January 2020, concerned the alleged use of gender quotas in entrance tests for physical education teaching. The Ombudsman found an infringement of the Act on Equality.

Stimulating girls' interest in STEM

Most often, programmes and initiatives aim to increase female participation in STEM education and/or are intended to help deconstruct gender prejudices and stereotypes in pupils' and students' educational and vocational choices. They often contain promotional activities and campaigns specifically focused on stimulating girls' interest in technology and IT. On an international level, for example, global private companies such as Microsoft have launched initiatives like DigiGirlz, which aims to 'give middle and high school girls opportunities to learn about careers in technology'.⁶⁹⁵

Similar programmes and initiatives are available in many European countries. In **Austria**, a 'Girls' day' has been introduced, the aim of which is 'to encourage girls to choose professions outside the traditional role models'.⁶⁹⁶ In **Belgium**, every year, the French-speaking Community organises a 'Girls' Day, Boys' Day' in cooperation with the Institute for the Equality of Women and Men.⁶⁹⁷ The **Estonian** NGO Tech Sisters has an event series, Digigirls, that aims to introduce technology and IT career options to 7-12 year-old girls by bringing female IT professionals into the classroom to talk about their work and experience.⁶⁹⁸ Similarly, the interest of 7-14 year-old girls in technology, robotics and science is stimulated by offering free workshops and summer camps, supported by the private company, HK Unicorn Squad.⁶⁹⁹ In **Greece**, the Ministry of Education has announced the introduction of pilot 'dexterity labs' in school curricula from September 2020 onwards in order to familiarise pupils with new technologies and promote digital skills, notably among girls.⁷⁰⁰ Private companies such as Vodafone have raised awareness in Greece through campaigns such as 'STEM for girls – STEM education is not a boy's privilege. Girls let's change the world together',⁷⁰¹ and 'Learning to write a code for girls #codelikeagirl'.⁷⁰² Educational initiatives in Greece include campaigns such as the series of interactive workshops on 'STEM: Science for... girls' organised by the Hrakleidon Museum for 11-15 year-old female students in the context of a collaboration with the international NGO Greenlight for Girls (g4g).⁷⁰³ The **Hungarian** Women in Science Association has been running a 'Girls' Day' career orientation programme for girls aged 10-18 since 2012.⁷⁰⁴ In Hungary, moreover, private actors are involved in offering IT trainings to 9th and 10th grade girls (aged 14-16) as part of the Smartiz programme.⁷⁰⁵ In **Malta**, the eSkills Foundation has published guidelines on how to increase and retain women in ICT, which include recommendations on 'encouraging girls to consider a career in ICT' through changing parents' and students' perceptions and familiarising young women with ICT.⁷⁰⁶ One of the initiatives taken in this respect in the **Netherlands** is the 'technology pact', funded by the Ministry of Education, Culture and Welfare, in which companies, educational institutions and the Government cooperate to improve the connection between tech studies and the labour market.⁷⁰⁷ The ministry also funds the awareness-raising event 'Girlsday', which aims to bring girls into contact with technical organisations and companies.⁷⁰⁸ In **Poland**, targeted initiatives have been initiated by

695 DigiGirlz, see <https://www.microsoft.com/en-us/diversity/programs/digigirlz/default.aspx>.

696 See <https://www.bmbwf.gv.at/Themen/HS-Uni/Gleichstellung-und-Diversität/Policy-und-Maßnahmen/Förderung-von-Frauen-im-MINT-Bereich.html>.

697 See <http://www.gdbd.be>. Similarly, the Platform 'Women in tech', supported by the Brussels Capital Region, aims *inter alia* to 'sensitize and inform women about digital opportunities and to promote gender equality in the high-tech industry', for instance through 'rais[ing] awareness among young girls of STEM studies and careers' and submitting 'new gender policies to the institutional and governmental authorities with the contribution of [...] partners' expertise'; see <https://www.womenintech.brussels>.

698 See www.ituudised.ee/uudised/2018/10/09/tech-sisters-toi-kokku-koolitudrukud; <https://www.facebook.com/TechSisters/>.

699 See <https://unicornsquad.ee/2020/05/11/digigeenius-taavi-kotka-robootikakool-tudrukutele-on-nii-populaarne-et-see-tehakse-koigile-tasuta-kattesaadavaks/> and <https://unicornsquad.ee/unicornide-liikumine/>.

700 Kerameos, N (2020), 'Γυναίκες στις θετικές επιστήμες: Ένα στοίχημα που πρέπει να κερδίσουμε' (Women in science: A bet that we have to win) available at: <https://www.tovima.gr/2020/03/10/vimagazino/kerameos-gynaikes-stis-thetikes-epistimes-ena-stoixima-pou-prepei-na-kerdisoume/>.

701 See www.vodafonegenerationnext.gr/blog/stem-gia-koritsia.

702 See www.vodafone.gr/vodafone-ellados/digital-press-office/deltia-typou/20190621-ekmathisi-syggrafis-kodika-gia-koritsia-codelikeagirl-apo-ti-vodafone/.

703 See www.paidorama.com/stem-epistimi-gia-koritsia.

704 See <http://lanyoknapja.hu/>.

705 See <http://nokatud.hu/smartiz-jelentkezes/>.

706 See <https://eskills.org.mt/en/womeninict/Pages/Guidelines-to-Increase-and-Retain-Women-in-ICT.aspx>.

707 *Techniekpact* see www.techniekpact.nl/.

708 VHTO, Girlsday, see www.vhto.nl/projecten/girlsday.

civil society organisations in partnership with universities and public authorities. The NGO Perspektywy Education Foundation and the Conference of Rectors of Polish Technical Universities have led several national campaigns over the past 11 years promoting the participation of women in STEM education, including the 'Girls as Engineers' and 'Girls go Science!' campaigns, in cooperation with Government ministers and the Plenipotentiary for Equal Treatment.⁷⁰⁹ The Foundation describes these campaigns as successful and asserts that the number of girls choosing STEM education increased from 29 % to 37 % in Poland during this period.⁷¹⁰ In **Spain**, some universities have launched their own promotional initiatives, such as *INSPIRA*, *Mujeres con Ciencia*, *Mujer e Ingeniería*, a programme to attract young women to ICT careers. The Spanish Institute for Women has launched two mentoring and funding programmes: 'Diana', which offers mentoring to young girls in scientific studies and 'AHORA TU', which aims to increase the representation of women in IT studies through scholarships.⁷¹¹

Financial incentives and dedicated educational programmes

In addition to such promotional and awareness-raising activities, in a few countries, participation of girls and women in STEM education is further stimulated by financial incentives, by enhancing the gender inclusiveness of existing tech studies, by introducing dedicated educational programmes, or by specifically recruiting women to participate in tech studies. In **Austria**, for example, a policy instrument provides for dedicated funding for female students in tech studies, in collaboration with private companies.⁷¹² A similar example can be found in **Belgium**, where the Wallonian region has used the European Social Fund to co-finance a project to stimulate more women (particularly, women who are unemployed) to work in the ICT sector.⁷¹³ The **French** President, in a discourse on artificial intelligence in 2018, declared that in this field, 'there are many white males in their forties trained mainly in the big American or European universities. So we need to mobilise more scientific training in order to attract more women, to bring about greater equality and social diversity. I think that in the challenge we are taking up, there is a need for inclusiveness in these training programmes'.⁷¹⁴ However, it has been noted by the expert for France that the implementation of such measures has been slow.⁷¹⁵ More tangibly, a measure aimed at the integration of women in digital education has been put in place by the University of **Liechtenstein**, the Economic University of Vienna (**Austria**) and the University of Economics and Law of Berlin (**Germany**). This international project is entitled 'Gender Equality in Digital Entrepreneurship' and encompasses the development of an interdisciplinary and gender-sensitive masters programme to promote gender equality in digital entrepreneurship.⁷¹⁶ Targeted recruitment measures have been put in place in **Norway** by Oslo University through a project called '*Jenter og informatikk*' (Girls and IT), which aims to facilitate the recruitment of girls in IT studies in higher education. The initiative includes various measures to increase the recruitment of women such as hosting an 'IT camp' for girls at high schools in Norway, organising a 'day for girls' for new undergraduate students at the IT faculty at the University of Oslo and lunches for new female masters students in IT at the university.⁷¹⁷ Similar initiatives have been put in place by the Norwegian University of Technology and Science, such as the *Jenteprosjektet* (Girls in technology)

709 See www.dziewczynynapolitechniki.pl/english.

710 See www.dziewczynynapolitechniki.pl/english.

711 These and other initiatives are listed in the *White Book of Women in Technology*, available at: www.mineco.gob.es/stfls/mineco/ministerio/ficheros/libreria/LibroBlancoFINAL.pdf, 98-100.

712 See <https://www.bmbwf.gv.at/Themen/HS-Uni/Gleichstellung-und-Diversität/Policy-und-Maßnahmen/Förderung-von-Frauen-im-MINT-Bereich.html>.

713 See www.interface3.be or see (presented in English as a good practice) Annex 2, 29 of the document available at: <https://ec.europa.eu/esf/transnationality/content/new-dossier-female-employment>. See also Wuiame N (2018), 'Female (un)employment and worklife balance' (Employment Network, Technical dossier No. 8, Transnational ESF Platform).

714 President Macron promised to follow up the proposals of the Villani Report in a speech at the Collège de France in 2018, see www.elysee.fr/emmanuel-macron/2018/03/29/discours-du-president-de-la-republique-sur-lintelligence-artificielle. See www.elysee.fr/emmanuel-macron/2019/10/30/cloture-du-global-forum-on-artificial-intelligence-for-humanity-par-le-president-emmanuel-macron.

715 See www.lemonde.fr/pixels/article/2018/07/30/intelligence-artificielle-la-diversite-humaine-un-enjeu-central-pour-son-developpement_5337428_4408996.html. See also Devillers L (2017), *Des robots et des hommes* (On robots and humans) (Plon).

716 See further details on www.gender-digitalent.com.

717 See the University's website: www.mn.uio.no/ifi/studier/jenter-og-informatikk/.

project.⁷¹⁸ In **Poland**, the Perspektywy Education Foundation launched the ‘Girls Go Start-up! Academy’ in cooperation with the Association TOP500-Innovators, which is an academy of innovation for female students and graduates of STEM studies.⁷¹⁹ In addition to the activities of the Foundation, scholarship and mentoring programmes targeted at women such as the new technologies for girls and the IT for SHE programmes have been put in place.⁷²⁰ The **Slovakian** NGO, You too in IT, established in 2012, aims to achieve a female participation of more than 30 % in IT curricula and over 40 % in IT professions as well as to change gender stereotypes in ICT sectors.⁷²¹ To do so, for instance, it organises a Women Tester Academy, which allows women to train as certified software testers and thus increases their chances of starting a professional career in IT.⁷²²

4.3.3.3 Efforts to increase gender diversity in employment in STEM fields

Targets and quotas

Only few countries have introduced positive action and quota measures to increase the number of women working in the tech sector. In 2018, a report commissioned by the **French** public authorities recommended the adoption of measures to diversify cohorts of software engineers and developers, with a target of 40 % female students in digital training by 2020.⁷²³ At other times, targets are much less clearly formulated. In the **United Kingdom**, for example, the independent review ‘Growing the Artificial Intelligence Industry in the UK’, published in 2017, highlighted the importance of a diverse workforce to ‘ensure that algorithmic biases are avoided in the selection of training data, design of algorithms and networks and the delivery of products and services’ and made a number of recommendations to that end.⁷²⁴ In response, the UK Government’s policy paper ‘AI Sector Deal’, published in 2019, includes a commitment to working with the AI Council (an independent expert committee) to increase diversity in the AI research base and workforce, but no specific targets for the representation of women in the technology sector have been formulated so far.⁷²⁵

This clearly shows that there are hardly any binding positive action measures on the side of public authorities in Europe. Indeed, policy-making bodies appear to be wary of such positive action, in particular quota policies, although such temporary measures could dramatically contribute to closing the gender digital gap in the future. In **Malta**, for example, a study analysing the gender gap in the digital sector reported opposition towards positive action.⁷²⁶

Promotional programmes and training courses

Much more common than targeted positive action measures or quotas are programmes and initiatives to increase female representation in the IT sector, which are often run by private (tech) companies or civil

718 See www.ntnu.no/jenterogteknologi/jenteprosjektet.

719 See www.dziewczynynapolitechniki.pl/english.

720 For ‘IT for SHE’, see: www.itforshe.pl/pl/nagroda-2017 and <http://www.itforshe.pl/pl/#o-programie>; for ‘New technologies for girls’ see: www.stypendiadladziewczyn.pl/en and <https://www.stypendiadladziewczyn.pl/pl/aktualnosci/66-to-juz-szosta-edycja-programu-stypendialnego-dla-studentek-politechnik>.

721 See www.ajtyvit.sk/.

722 See www.ajtyvit.sk/podujatia/women-tester-academy-iv.

723 Villani, C (2018), *Donner un sens à l’intelligence artificielle : pour une stratégie nationale et européenne* (Giving meaning to artificial intelligence: for a national and European strategy) Report for the Prime Minister, (8 September 2018) available at: www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf.

724 Hall, W and Pesenti, J (2017), *Growing the Artificial Intelligence Industry in the UK*, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf, 56.

725 See www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal. On the follow-up, see <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/1544/154402.htm> and BEIS (2018), ‘Government response to House of Lords Artificial Intelligence Select Committee’s Report on AI in the UK: Ready, Willing and Able?’ (CM 9645, June 2018) available at: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>.

726 ‘[M] ost respondents were against forms of positive discrimination’, see eSkills Foundation, ‘Analysis of the Gender gap in the Digital Sector in Malta’, available at: https://eskills.org.mt/en/analysis%20of%20the%20gender%20gap%20in%20the%20digital%20sector/Documents/Analysis_of_the_gender_gap_in_Maltav2.pdf, 5.

society, supported by or in partnership with Government bodies. Many such programmes are promotional or supportive in nature, for instance providing for additional on-the-job training courses and workshops. In **Estonia**, for example, the Ministry of Economic Affairs and Communications has launched a programme called StartIT, which aims to encourage girls and young women to study and work in the field of ICT, by organising workshops in technology companies.⁷²⁷ Similar initiatives exist in **Germany**, which include the Ada-Lovelace-project, the mission of which is to 'break up traditional role models and strengthen the positive image of STEM professions' among women,⁷²⁸ and collaborative projects between tech industry, science and young professionals such as GEWINN (*Gender/Wissen/Informatik/Netzwerk*), which mobilises research on gender and computer science to support female young professionals in computer science aspiring to top positions.⁷²⁹ In **Hungary**, *Techcsajok* (TechGals) offers training and awareness-raising events of job opportunities in tech for young women and girls.⁷³⁰ In addition, a special department of Women in Education, Research and Informatics was created in 2019 within the John von Neumann Computer Society to promote the role of women in IT.⁷³¹ Since the tech sector in **Latvia** faces a shortage of workers, in 2019, the IT company Accenture Latvia launched an educational programme to offer programming training and scholarships to 160 women per year.⁷³² In **Luxembourg**, the NGO Women in Digital Empowerment (WIDE) was established in 2014 to help build a more diverse pool of ICT talent in Luxembourg by organising conferences, training courses and workshops promoting the participation of women in IT and STEM-related professions.⁷³³ Examples of such initiatives include the startup leadership programme, which assists women who want to create an IT business. Civil society has also engaged with the issue in Luxembourg. For example, the Rails Girls in Luxembourg programme offers a free-of-charge introduction to computer programming for teachers, students and women re-entering the labour market in the marketing and communication sector. The NGO also organises free introduction to coding courses.⁷³⁴ The **Slovakian** Government has launched an action plan to help increase the share of women in the IT and digital sectors, for instance by offering support to projects that promote the engagement of women in IT jobs.⁷³⁵ The action plan also aims to increase cooperation with the private sector to organise stays, study visits and workshops for women in IT companies.⁷³⁶

Awareness-raising among employers and recruiters

In addition, some initiatives focus on raising awareness among recruiters, employers and other stakeholders about the need to ensure gender diversity in STEM fields. In **France**, for example, when it comes to IT professions, the National Digital Council has published a charter of good practice for more diversity in companies, which encourages signatories to pay attention to the gender balance in recruitment and funding.⁷³⁷ The **German** project, gendering MINT digital, offers lectures on gender and IT, gender equality in STEM and gender in techno-scientific literacy.⁷³⁸ The **Maltese** eSkills Foundation has published guidelines for increasing the number of women in technology professions that include paying attention to the non-discriminatory wording of job ads, introducing equality targets and ensuring diverse interview panels.⁷³⁹ Similarly, in **Poland**, the Perspektywy Education Foundation has published a set of

727 See www.workinestonia.com/state-started-campaign-inviting-young-women-study/; <https://startit.ee>.

728 See <https://ada-lovelace.de/>.

729 See www.kompetenzz.de/aktivitaeten/gewinn.

730 See www.techcsajok.hu/.

731 See https://njszt.hu/hu/szakmai_kozosseg/nokit.

732 BiSmart (2020) 'Motivētākās sievietes aicina pieteikties bezmaksas apmācībām IT jomā' (The most motivated women are invited to apply for free of charge studies in IT sector) (25 April 2020), available at: <https://bismart.lv/blogs/vieda-valsts/sievietes-acinatas-pieteikties-bezmaksas-programmesanas-apmacibam-1420>.

733 See <https://wide.lu/>.

734 See <https://wide.lu/> and www.cc.lu/uploads/tx_userccpublications/COVER_STORY_OK.pdf.

735 The action plan was adopted by Government Resolution No 337/2019 of 3 July 2019, available at: www.vicpremier.gov.sk/wp-content/uploads/2019/10/AP-DT-English-Version-FINAL.pdf, 26.

736 See Government Resolution No 337/2019 of 3 July 2019, available at: www.vicpremier.gov.sk/wp-content/uploads/2019/10/AP-DT-English-Version-FINAL.pdf, 26.

737 See <https://cnnumerique.fr/nos-travaux/acceleration-du-financement-des-femmes-entrepreneures>.

738 See www2.hu-berlin.de/genderingmintdigital/.

739 eSkills Foundation, 'Analysis of the Gender gap in the Digital Sector in Malta', available at: https://eskills.org.mt/en/analysis%20of%20the%20gender%20gap%20in%20the%20digital%20sector/Documents/Analysis_of_the_gender_gap_in_Maltav2.pdf, 4.

recommendations to increase the participation of women in the tech industry addressed to educational institutions, tech companies, the state, NGOs and the media.⁷⁴⁰ The Alan Turing Institute, which is the national institute for data science and artificial intelligence in the **United Kingdom**, has established a Women in Data Science and AI Hub, which aims to inform policymakers about increasing the number of women in data science and AI through research.⁷⁴¹

Employment conditions and culture

Some other initiatives focus on employment conditions and workplace culture in IT professions. In recent years, for instance, there has been increasing awareness in **Estonia** that women's participation in tech should be supported by family-friendly policies and a women-friendly company culture. In this context, some private companies such as Skype Technologies OÜ, Fortumo OÜ and Twilio Estonia OÜ have proposed flexible working time arrangements, remote working and flexible leave policies to support female employees.⁷⁴² In **Malta**, the eSkills Foundation has published a number of guidelines to help increase the recruitment of women in ICT, which include 'offering a women-friendly work environment'.⁷⁴³ The guidelines also include recommendations on retaining the female workforce in ICT jobs and female leadership.

Peer support, networking and mentoring

A focus on strengthening peer interaction between women, mentoring and building support networks is also a common part of many programmes and activities in Europe. In **Germany**, for example, career networks such as FEMTEC offer female tech professionals training, industrial contacts and networking opportunities.⁷⁴⁴ Among existing mentorship and counselling programmes in **Lithuania**, 'women go tech' addresses women seeking professional careers in the technology sector. The programme was launched in 2016 to strengthen gender equality and female leadership in the labour market. The main goal is to create '500 women in tech success stories' by 2021 by helping women to enter the tech labour market, obtain promotions in this sector and establish a tech start-up.⁷⁴⁵ The Lithuanian national initiative 'Women & Technology' aims to connect women working in the IT sector, enhance their skills and knowledge, allow them to share their experience, and to promote the acquisition of IT skills among the female youth.⁷⁴⁶ Another example of this is the set of activities proposed by the **Norwegian** organisation Oda Network, which cooperates with private companies to offer networking opportunities, mentorship programmes, awards and awareness-raising events aimed at enhancing the participation and visibility of women in tech.⁷⁴⁷ In a similar vein, the Perspektywy Education Foundation in **Poland** has initiated a project called Lean in STEM!, which aims to promote 'technical and scientific education and careers in the technological industry and other STEM-related areas among young women' through 'supporting [the] creation of a female networking culture in the technological industry and the STEM area'. In partnership with private IT and tech companies, it offers a mentoring programme, virtual meetings, 'technological teatimes' and offline meetings in technological companies as well as holding a conference entitled 'Women in

740 See <http://potencjalkobiet.pl/recommendations>.

741 See www.turing.ac.uk/about-us/equality-diversity-and-inclusion/women-data-science-and-ai.

742 Compare Lindma, LM (2019), 'Kaasamise ja mitmekesisuse kiituseks' (In praise of inclusion and diversity), *Director* 5/2019, available at: <https://director.ee/2019/05/02/kaasamise-ja-mitmekesisuse-kiituseks/>; see further e.g. www.twilio.com/company/diversity and <https://fortumo.com/careers/home/culture/>. See also Laas, A (2017), *Practices with Potential* (unpublished national report of Estonia for EIGE study on work-life balance in the ICT sector) available at: <https://eige.europa.eu/publications/work-life-balance/work-life-balance-in-ict>). More information on the Skype research project from 2015 can be found with Kindsiko, E, Türk, K and Kantšukov, M (2015), *Naiste roll ja selle suurendamise võimalused Eesti IKT sektoris* (Women's role and potential in ICT sector); summary available in English by Külmoja, I (2015), 'Six Myths Keeping Estonian Girls Out of IT', available at: <http://blog.ut.ee/six-myths-keeping-estonian-girls-out-of-it/>.

743 eSkills Foundation, 'Analysis of the Gender gap in the Digital Sector in Malta', available at: https://eskills.org.mt/en/analysis%20of%20the%20gender%20gap%20in%20the%20digital%20sector/Documents/Analysis_of_the_gender_gap_in_Malta2.pdf.

744 See www.femtec.org/.

745 See www.womengotech.lt/en/.

746 See www.mitt.lt/?q=en.

747 See <https://odanettverk.no/>.

Tech Summit 2018'.⁷⁴⁸ Similarly, in **Slovakia**, the NGO You Too in IT, established in 2012, organises mentoring opportunities.⁷⁴⁹ In **Sweden**, since 2006, the employers' organisation in the IT and telecom industry, Almega, has run a leadership programme called Womentor to support companies in their efforts to increase the proportion of women in management positions. It has been reported that, on average, the share of women who have accessed management positions in companies that have participated in Womentor increased from 25 % to 34 % between 2006 and 2016, whereas the average share of women decreased from 32 % to 28 % for the industry as a whole during the same period.⁷⁵⁰ It should also be noted here that there are also global programmes and initiatives with a similar objective. An example is the international support network, women in data science, which aims to create a forum for women to enhance their network, knowledge and cooperation opportunities within the data science field.⁷⁵¹

EU efforts

Finally, in addition to the national efforts discussed above, it can be noted that the EU Member States and associated countries have embarked on common efforts to solve the existing problems of underrepresentation of women in STEM. In 2019, **Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia Slovenia, Spain, Sweden** and the **United Kingdom** signed the Women in Digital Declaration, committing to 'work closely with the public and private sectors and civil society to achieve equality in tech'.⁷⁵² The objectives of the declaration include creating a national strategy to encourage women's participation in the digital sector, establishing a European girls and women in ICT day, encouraging the media to promote positive representations of women in digital sectors, inciting companies to act against gender discrimination in the workplace, advancing more gender-balanced company boards and improving both the monitoring of the problem and data collection on the participation of women in the digital field.

4.3.3.4 Impact of measures and initiatives

The sections above have exposed a flurry of promotional measures, empowerment initiatives and support programmes, organised by public and private players as well as civil society organisations, both nationally and on an EU and even a global level. These activities appear mainly to be aimed at familiarising female students with job opportunities in AI and data science (and STEM more broadly), at deconstructing existing gender stereotypes so as to combat gender segregation in professional communities involved in AI research and development, and at offering female professionals networking opportunities. Several studies show that such measures are both effective and important in tackling the gender digital gap, and they can indeed be regarded as good practice. Research, for example, demonstrates that eliminating gender stereotypes and raising the profile of female role models and mentors is crucial to closing the digital gender gap.⁷⁵³ Promotional and awareness-raising activities directed at women also play an essential role in increasing the rate of women in IT education and professions.⁷⁵⁴

However, the long list of activities also shows that recognition of the issue of the lack of diversity in AI remains limited to gender inequality without addressing other dimensions of discrimination, such as the absence of ethnic minorities or workers with disabilities in professional communities. For example,

748 See www.dziewczynynapolitechniki.pl/english.

749 See www.ajtyvit.sk/.

750 See www.womentor.se/.

751 See www.widsconference.org/.

752 See European Commission (2019) 'EU countries commit to boost participation of women in digital' (9 April 2019), available at: <https://ec.europa.eu/digital-single-market/en/news/eu-countries-commit-boost-participation-women-digital>.

753 Kamberidou, I, and Pascall, N (2020), 'The digital skills crisis: engendering technology—empowering women in cyberspace', 4 *European Journal of Social Sciences Studies* (6) 1, available at: www.tinyurl.com/ulxI9od.

754 Fundacja Edukacyjna Perspektywy (2019) *Women at polytechnics (Kobiety na politechnikach)* available at: www.dziewczynynapolitechniki.pl/pdf/raport-kobiety-na-politechnikach2019.pdf. The report is based on the publication of the Central Statistical Office of Poland, *Szkoły wyższe i ich finanse* (Higher education institutions and their finances).

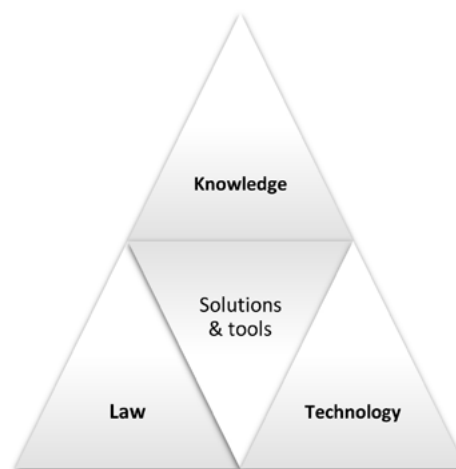
in **Bulgaria**, the national expert has reported that, despite a general commitment to enhance the participation of women in technology in the National Gender Equality Strategy 2016-2020, specific measures are insufficient and, in addition, inclusion plans in STEM education do not extend to ethnic minorities such as the Roma.⁷⁵⁵ The **Croatian** expert also reports that there is no information available on potential problems relating to the involvement of other minority groups in ICT or data science education and professions. This lack of attention to the underrepresentation of minority groups in AI is problematic because it means that the perspectives, needs and experiences of entire parts of society are *de facto* ignored in AI design and development. This may lead to algorithmic applications that are sub-optimally calibrated for these minorities and that bear the risk of discriminating against them.

Finally, it should be noted that although it is certain that '[i]ncreasing women's [and other minorities'] participation is the only way to ensure that their perspectives and priorities will inform the insights that this new field will generate, and the uses to which it is put in society',⁷⁵⁶ it is not enough to avoid biases from causing unlawful discrimination in AI design. In addition to being members of an increasingly diverse and balanced community, AI professionals and data scientists need to be specifically trained to recognise, avoid and test for these biases when designing algorithmic applications. One way to conduct such training would be by adapting university curricula and vocational and professional training to include digital humanities, social sciences and ethics components so as to train graduates to recognise these biases and put debiasing strategies in place in order to avoid risks of algorithmic discrimination. These measures should be complementary to those listed above in relation to equal participation in professional communities.

4.4 Potential solutions and tools to prevent and remedy algorithmic discrimination: a tridimensional approach

4.4.1 Introduction

It is clear from the long list of measures, initiatives, policy guidelines, codes of conduct and other activities covered in section 4.3 that there are many examples of good practice in Europe to help meet the challenges and counter the risks and problems related to algorithmic decision making. In addition, section 4.2 shows that, in addition to such challenges, problems and risks, the use of algorithms can bring considerable benefits and opportunities. Bringing the findings of these sections together, and combining them with the *theoretical* insights discussed in Chapters 1 and 2, this section offers a systematic review of existing and novel solutions and tools to tackle algorithmic discrimination. These are numerous and can be classified under three main categories: *legal*, *knowledge-based* and *technology-based* solutions and tools:



755 The topic is absent from the *National Strategy for integration of Roma 2012-2020*, see https://ec.europa.eu/info/sites/info/files/roma_bulgaria_strategy_en.pdf.

756 See www.turing.ac.uk/about-us/equality-diversity-and-inclusion/women-data-science-and-ai.

Section 4.4.2 identifies a number of legal solutions, relating to both the EU's non-discrimination legislation and the interpretation and application of specific non-discrimination concepts by the Court of Justice of the EU. Subsequently, in section 4.4.3, a number of tools and instruments are discussed that relate to knowledge building, awareness raising and training. Section 4.4.4 then turns to addressing the need for technology-based solutions to the problems of algorithmic discrimination, such as *ex ante* strategies of 'non-discrimination by design' and *ex post* strategies of creating accountability mechanisms. Finally, we argue that it is not enough to focus on just one or two of the three sets of available strategies to counteract algorithmic discrimination and make the best possible use of the existing benefits and opportunities. Instead, it is essential to combine all three sets of tools and instruments, using a holistic approach. In the final section of this chapter, section 4.5, we make a proposal for taking such an approach using the acronym PROTECT.

4.4.2 Legal solutions

The law has a key role to play in combating algorithmic discrimination. While ethical solutions should not be neglected, a hard law approach will be necessary to effectively meet this challenge. Various solutions exist, with varying degrees of intervention, and can be considered together. In view of requirements for technology neutrality, we do not recommend the creation of a new body of laws specifically tackling algorithmic discrimination, but rather the targeted adaptation and purposive interpretation of existing legislation, doctrinal practices and institutional arrangements. At the EU level, solutions range from reforming the legislation to revisiting existing concepts, adapting the doctrine and adjusting institutional arrangements.

4.4.2.1 Taking a more flexible approach to protected grounds and intersectionality

First, as was shown in section 2.2, taking a more flexible approach to protected grounds is necessary in order to be able to effectively tackle algorithmic discrimination, in particular when it manifests as proxy or as intersectional discrimination. In this regard, the open-ended list of protected grounds listed in Article 21 of the Charter of Fundamental Rights, taken in combination with the general principle of non-discrimination, could play a particularly important role.⁷⁵⁷ On the one hand, the non-exhaustive nature of the provision could help judges tackle algorithmic discrimination by contextually defining 'new' protected grounds in order to expand the non-discrimination protection to proxy-based or intersectional forms of algorithmic discrimination. By analogy to the European Court of Human Rights, which sometimes relies on the mention of 'other status' in order to contextually protect grounds of discrimination such as health status,⁷⁵⁸ that are not listed in Article 14 of the European Convention on Human Rights, the Court of Justice could rely on the open-ended formulation of Article 21 of the Charter to accommodate proxy-based or intersectional discrimination by treating these cases as analogous to discrimination based on protected grounds.⁷⁵⁹ In practice, if a proxy does not fall neatly within the scope of a protected ground, or if the basis for an algorithmic recommendation is composite and only disadvantages a subgroup within a protected group, contextually recognising an analogous protected ground could help perform the usual comparator-based test and capture algorithmic discrimination.

On the other hand, Article 21 of the Charter could be used to fill some of the gaps arising from the 'hierarchy of grounds' that structures EU non-discrimination legislation.⁷⁶⁰ Indeed, where Member States are implementing EU law in line with Article 51 of the Charter, but a given matter falls outside the

757 See further Gerards, JH and Zuiderveen Borgesius, F (forthcoming), 'Protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence' (in preparation, on file with the authors).

758 See *Kiyutin v. Russia*, Application no. 2700/10 (2011).

759 See also the approach of the Canadian Supreme Court in treating intersectional grounds of discrimination as 'analogous', e.g. Aylward, C, (2010) 'Intersectionality: Crossing the Theoretical and Praxis Divide' 1 *Journal of Critical Race Theory* 1 and Grabham, E (2002), 'Law v Canada: New Directions for Equality under the Canadian Charter?' 22 *Oxford Journal of Legal Studies* 641.

760 See Kilpatrick, C (2014), 'Article 21- Non-Discrimination' in Peers, S and others (eds), *The EU Charter of Fundamental Rights A Commentary* (Hart Publishing 2014).

material scope of the non-discrimination directives, the Charter could play a subsidiary role to guarantee the protection of equality rights.⁷⁶¹ The Court has validated this approach in cases such as *Léger*, where rules on blood donation excluded men who have same-sex sexual relations, a matter which fell outside the material scope of Directive 2000/78/EC.⁷⁶² However, the Court has interpreted the potential of the Charter in this regard in a rather restrictive way so far.⁷⁶³ In addition, the Court has categorically ruled out the possibility of expanding the list of protected grounds in matters falling within the scope of the equality directives in cases such as *Chacon Navas* and *Kaltoft*.⁷⁶⁴ By contrast, however, cases such as *Egenberger* and *IR* might be read as supporting a wider interpretation in line with the general principle of non-discrimination.⁷⁶⁵ Hence, Article 21 of the Charter could offer valuable remedies to algorithmic discrimination, and its potential can be more fully explored.

In addition, intersectional forms of discrimination have been shown to be particularly pervasive in algorithmic decision making, which often rely on proxies or grounds positioned on the interface between different protected grounds of discrimination. Such forms of discrimination could be captured by relying on the concept of ‘multiple discrimination’ established in Recital 14 of the Racial Equality Directive and Recital 3 of the Equal Treatment Directive. Although this particular notion seems to be limited to gendered forms of intersectional discrimination – ‘especially [...] women are often the victims of multiple discrimination’ – it is not framed in an exhaustive way and could be extended to other protected grounds of discrimination. Although the Court of Justice rejected a finding of intersectional discrimination in *Parris*, AG Kokott has suggested that recognising the concept of intersectionality could enrich the Court’s assessments of discrimination.⁷⁶⁶ Moreover, the Court has implicitly recognised intersectional discrimination in some instances, showing that EU law does offer some opportunities to combat intersectional forms of algorithmic discrimination. For example, in its decisions in *Odar* and *Bedi*, the Court has implicitly recognised how disadvantage arises from the interaction of age and disability based discrimination, acknowledging ‘the risk that severely disabled persons may have financial requirements arising from their disability which cannot be adjusted and/or that, with advancing age, those financial requirements may increase’.⁷⁶⁷ Finally, the latest Gender Equality Strategy 2020–2025, published by the European Commission in March 2020, recognises intersectionality as a ‘cross-cutting principle’ and indicates that ‘the intersectionality of gender with other grounds of discrimination will be addressed across EU policies’.⁷⁶⁸ This means that, at the very least, there are some useful starting points available to deal with the particular problems of proxy discrimination and intersectional discrimination that have been discussed in section 2.2.

-
- 761 See Bribosia, E, Rorive, I and Hilaire, J (2017), ‘Article 21: Non Discrimination’ in Picod, F, Van Drooghenbroeck, S and Rizcallah, C (eds), *Charte des droits fondamentaux de l’Union européenne: Commentaire article par article* (Charter of Fundamental Rights of the European Union: Article-by-article commentary) (Brussels, Bruylant). See also Gerards, JH and Zuiderveen Borgesius, F (forthcoming), ‘protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence’ (in preparation, on file with the authors).
- 762 See Judgment of 29 April 2015, *Geoffrey Léger v Ministre des Affaires sociales, de la Santé et des Droits des femmes and Etablissement français du sang* C-528/13 EU:C:2015:288.
- 763 See Ward, A (2018), ‘The Impact of the EU Charter of Fundamental Rights on Anti-Discrimination Law: More a whimper than a bang?’ 20 *Cambridge Yearbook of European Legal Studies* 32.
- 764 Judgment of 11 July 2006, *Sonia Chacón Navas v Eurest Colectividades SA* C-13/05 EU:C:2006:456 and Judgment of 18 December 2014, *Fag og Arbejde (FOA) v Kommunernes Landsforening (KL)* C-354/13 EU:C:2014:2463.
- 765 Judgment of 17 April 2018, *Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung eV* C-414/16 EU:C:2018:257 and Judgment of 11 September 2018, *IR v JQ* C-68/17 EU:C:2018:696.
- 766 Judgment of 24 November 2016, *David L. Parris v. Trinity College Dublin and Others*, C-443/15, EU:C:2016:897 and Opinion of Advocate General Kokott, *David L. Parris v Trinity College Dublin and Others*, C-443/15, EU:C:2016:493, 2016.
- 767 Judgment of 6 December 2012, *Johann Odar v Baxter Deutschland GmbH*, C-152/11, EU:C:2012:772, [69] and Judgment of 19 September 2018, *Surjit Singh Bedi v Bundesrepublik Deutschland and Bundesrepublik Deutschland in Prozessstandschaft für das Vereinigte Königreich von Großbritannien und Nordirland*, C-312/17, EU:C:2018:734, [75]. See Xenidis, R (2020), ‘Tuning EU equality law to algorithmic discrimination: Three pathways to resilience’, p. 744 and for an analysis of the ‘intra-categorical approach’ to intersectional discrimination adopted by the Court, see Xenidis, R (2020) *Beyond the ‘Master’s Tools’: Putting Intersectionality to Work in European Non-Discrimination Law. A Study of the European Union and the Council of Europe Non-Discrimination Law Regimes* (Florence: European University Institute).
- 768 European Commission (2020), ‘A Union of Equality: Gender Equality Strategy 2020–2025’, 2 and 16.

4.4.2.2 Widening the material scope of the gender equality and non-discrimination directives

Besides giving a more flexible interpretation to the Charter, extending the material scope of the gender equality and non-discrimination directives would provide a more comprehensive basis to tackle algorithmic discrimination. As highlighted in previous Chapters of this report, the 'hierarchy of grounds' that characterises EU non-discrimination legislation is highly problematic. Indeed, algorithmic discrimination is likely to arise in areas where only race and gender equality are protected, and in particular in the market for goods and services. In this regard, the Horizontal Directive proposed by the Commission in 2008, which is still subject to negotiations in the Council, could provide an important response to the problems, risks and challenges highlighted in this chapter.⁷⁶⁹ Adopting this directive would extend the protection against discrimination on grounds of age, disability, sexual orientation and religious beliefs to the field of goods and services. That said, the gaps relating to the exceptions concerning the media, advertising and education in the Gender Goods and Services Directives 2000/4/EC would remain an issue, since EU law does not provide a legal basis to combat algorithmic discrimination in these areas in Member States where these fields have been allowed to remain outside the Directive's scope of protection.

4.4.2.3 Re-interpreting the concept of 'instruction to discriminate'

The concept of 'instruction to discriminate' could provide an interesting basis for a doctrinal innovation in relation to the classification and responsibility challenges delineated in Chapter 2.⁷⁷⁰ The EU gender equality and non-discrimination directives all include the concept and qualify it as a form of discrimination. The Equal Treatment Directive 2000/78/EC prohibits instructions to discriminate in its Article 2(4); the Gender Goods and Services Directive 2004/113/EC in Article 4(1); the Gender Equality Directive (recast) 2006/54/EC in Article 2(2)(b); and the Racial Equality Directive 2000/43/EC in Article 2(4).⁷⁷¹ While the concept is not defined in EU law and has not yet been subject to the Court's interpretation, an innovative interpretation of the notion of 'instruction to discriminate' could effectively mitigate the substantive and procedural hurdles that arise in the context of algorithmic discrimination.

First, at the conceptual level, the notion provides a good fit with the situation where discrimination arises from a decision supported by an algorithmic recommendation system. An algorithmic recommendation or prediction system can indeed be considered to produce an instruction about which decision to make in a given context. If such algorithmic output is discriminatory, it could then be considered an instruction to discriminate. It has been argued that the concept of instruction to discriminate is intended at capturing situations in which an employee is ordered by his or her superior to discriminate against protected groups, for instance if a real estate agency asked its employees not to rent flats to people with a certain ethnic background. The notion of instruction is intended to capture the liability of the employer in situations of delegated discrimination. Moreover, the concept of instruction to discriminate not only covers situations of coercion (e.g. an order), but also includes incitement to discriminate.⁷⁷² In situations of algorithmic discrimination, the 'instruction' that results from an algorithm can be understood as an incitement. It is not coercive and can be overridden by human decision makers, yet the automation bias discussed in section 1.4.1 means that such instructions will often be implemented.

Secondly, the concept of instruction to discriminate could avoid the uncertainties regarding the classification of algorithmic discrimination as direct or indirect discrimination. As explained in Chapter 2, the difficulties linked to the identification of differential treatment based on protected grounds in the context

769 European Commission (2008), *Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation* COM (2008) 0426 final.

770 This argument is further developed in Xenidis, R, 'Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law' (on file with the author).

771 More specifically, all these provisions set out that 'an instruction to discriminate' or an 'instruction to direct or indirect discrimination' on any one of the protected grounds 'shall be deemed to be discrimination within the meaning of [the EU anti-discrimination] Directive[s]'.
 772 See e.g. Busstra, M.J. (2011) *The implications of the Racial Equality Directive for minority protection within the European Union*. Eleven International Pub.

of algorithmic operations mean that the concept of indirect discrimination might become a conceptual 'refuge' to capture the discriminatory wrongs of algorithms. The concept of instruction to discriminate would instead allow the discriminatory output of algorithms to be treated not as discrimination in and of itself, but rather as a variable disadvantageously affecting protected groups in human decision making. Instead of placing the focus on the abstract algorithmic output, which casts doubts on the allocation of responsibility, the concept of instruction to discriminate highlights the role and responsibility of the human decision maker in discriminating against protected groups.

Thirdly, such a doctrinal classification of algorithmic discrimination would yield important procedural consequences regarding issues of responsibility and the need to open the algorithmic 'black box' of algorithmic decision making. In particular, the notion of instruction to discriminate could gain relevance as a tool to capture the liability of an employer or business using discriminatory algorithms. Such a conceptual innovation would favour active engagement as opposed to passive review by any humans in the loop, without the option of delegating responsibility to abstract entities, i.e. algorithms themselves. One could imagine that the designers of algorithmic models, to prevent any accusation of instructing to discriminate, would either self-assess their products for discrimination or would seek certification from specialised third-party agencies that their products are discrimination-free in order to sell them to users (public administrations, employers, businesses, etc). Similarly, users of algorithms would have to pay attention to such criteria and certifications when choosing algorithmic models in order to prevent accusations of discrimination. The management of liability could also be optimised by users of algorithms, for example through contractual clauses allowing them to seek damages with certification agencies to some extent if they are accused of discrimination in relation to a certified algorithm. By extension, this would foster the creation of an entire ecosystem designed to prevent discrimination. Such a clear allocation of responsibilities onto those who deploy algorithmic techniques could help prevent algorithmic discrimination *ex ante* and contribute to solving the responsibility challenge as described in 1.4.6.

4.4.2.4 Easing the burden of proof

Turning to the CJEU's non-discrimination doctrine, several interesting elements could provide ways to address the transparency and evidentiary problems exposed in sections 2.4 and 3.2.3. In the context of algorithmic discrimination, the burden of establishing a *prima facie* case that would trigger a shift of the burden of proof might indeed be too heavy for individual victims, in particular if no tests or audits are conducted, if statistics on the effects of given algorithms on given groups are unavailable, or if respondents refuse to disclose information on the effects of specific algorithms. In the event of such difficulties, the principle of effectiveness of EU law offers a possibility for national courts to consider the refusal to disclose information as contributing to the establishment of a *prima facie* case of discrimination, thus shifting the burden of proof to the respondent. In particular, the 'lack of transparency' doctrine established in *Danfoss* and the Court's approach in relation to the lack of access to information in *Meister* could assist victims in establishing *prima facie* evidence of algorithmic discrimination. In early equal pay cases, the CJEU made it clear that 'where an undertaking applies a system of pay which is totally lacking in transparency, it is for the employer to prove that his practice in the matter of wages is not discriminatory', meaning that the burden of proof shifts in order to safeguard the principle of effectiveness of EU law.⁷⁷³ The Court reasoned that doing otherwise would deprive potential victims of discrimination of any effective means to enforce gender equality law.⁷⁷⁴ Such an approach could help overcome situations where an algorithm is so complex (as might be the case for many machine-learning algorithms or interconnected algorithms) that its functioning is out of comprehension, or where trade secrets and IP rights protect an algorithmic model. Without obliging a company or employer to disclose the full model, the lack of transparency doctrine

⁷⁷³ Judgment of 17 October 1989, *Handels- og Kontorfunktionærernes Forbund I Danmark v Dansk Arbejdsgiverforening, acting on behalf of Danfoss* C-109/88 EU:C:1989:383 [16], confirmed in Judgment of 27 October 1993, *Dr. Pamela Mary Enderby v Frenchay Health Authority and Secretary of State for Health* C-127/92 EU:C:1993:859.

⁷⁷⁴ Judgment of 17 October 1989, *Handels- og Kontorfunktionærernes Forbund I Danmark v Dansk Arbejdsgiverforening, acting on behalf of Danfoss* C-109/88 EU:C:1989:383 [9].

would shift the burden of proof to the defendant, who would then be responsible for showing that there is no discrimination or that it can be justified in the context of the proportionality test.

In the context of equal pay, the recommendation on pay transparency published by the Commission could provide a further source of inspiration for transparency requirements in the context of algorithmic decision making.⁷⁷⁵ The Court's approach on the lack of access to information and the shift of the burden of proof was confirmed in *Kelly*, where the Court recognised that although applicants did not have a right to access relevant information, 'a refusal of disclosure by the defendant, in the context of establishing [a *prima facie* case of discrimination], could risk compromising the achievement of the objective pursued by [the] directive and thus depriving that provision in particular of its effectiveness'.⁷⁷⁶ In *Meister*, it went further by indicating that in such a situation, 'the fact that, [an] employer [...] seems to have refused [the applicant] any access to the information that she seeks to have disclosed' could be 'among the factors which may be taken into account' to establish a *prima facie* case of discrimination.⁷⁷⁷

4.4.2.5 Introducing a public and collective approach to monitoring and redress

At the institutional level, this report has also pointed out the obstacles that arise from the piecemeal approach of a redress system that is adversarial and based on individual litigation. Given the lack of transparency of algorithmic decision making and the ensuing difficulty in gathering evidence, as well as the costs of litigation in terms of time and financial resources, it is likely that individual litigants will be deterred from enforcing their own equality rights in courts. A solution to this problem could be to reinforce the supervisory role of public institutions, for example entrusting the monitoring and redress of algorithmic discrimination (and discrimination more broadly) to dedicated institutions such as equality bodies. This would imply reforming these institutions, providing them with adequate resources, investigation, auditing, supervision and sanctioning powers, the ability to render legally binding decisions (in the same manner as data protection authorities) as well as standing rights in courts (as is already the case in some EU Member States). At EU level, institutions such as the European Institute for Gender Equality (EIGE) could play an important role in the public monitoring of algorithmic discrimination by collecting, analysing and disseminating data on the impact on gender equality of algorithms used in the EU.

The practices discussed in sections 3.4.1.1 and 4.3.2 have also shown that there is potential (and need) for cooperation between equality bodies and data protection agencies in this regard. In addition, co-regulation mechanisms could also be put in place at EU and national level, and organisations with a legitimate interest could be entrusted with a more significant role in the enforcement of equality in the context of algorithms. In particular, legal standing could be generalised for organisations with a legitimate interest, such as trade unions and NGOs. Since algorithmic discrimination is by nature systemic, a collective approach would be more effective. In that sense, collective complaints mechanisms such as *actio popularis* and class action could provide interesting ways forward. Such institutional reforms could also secure a better enforcement of the existing legal framework as national experts have reported systematic deficiencies in the application of EU equality law by national courts.⁷⁷⁸

4.4.2.6 Accreditation, certification and supervision

The EU could further foster the creation of an accreditation system for certification and supervision in relation to algorithmic discrimination. For instance, one could imagine an EU accreditation scheme that would allow for the creation of certification agencies responsible for testing and licensing algorithmic models used in fields where discrimination is prohibited. The choice of institutional design (e.g. deciding whether these agencies are public or private) should then be left to the Member States. Principles such as

⁷⁷⁵ See European Commission Recommendation 2014/124/EU of 7 March 2014 on strengthening the principle of equal pay between men and women through transparency, OJ L 69, 8.3.2014.

⁷⁷⁶ Judgment of 21 July 2011, *Patrick Kelly v National University of Ireland (University College, Dublin)* C-104/10 EU:C:2011:506 [34].

⁷⁷⁷ Judgment of 19 April 2012, *Galina Meister v Speech Design Carrier Systems GmbH* C-415/10 EU:C:2012:217 [44], [47].

⁷⁷⁸ See section 3.4.2.2.

equality mainstreaming could play a role in certification systems, for instance as the basis for equality by design obligations for designers of algorithms seeking certification. These certification, accreditation and supervision systems could play an important role in realising the technology-based solutions, tools and instruments to be discussed in section 4.4.4.

4.4.2.7 Taking a multi-disciplinary legal approach

Finally, a multi-disciplinary legal approach integrating several fields of law (non-discrimination and gender equality law, data protection law, IP law, consumer protection law, etc.) and fostering cooperation between various legal players (e.g. equality bodies, data protection authorities with expertise in the field of new technologies, consumer associations, etc.) could prove key to adequately addressing algorithmic discrimination. As underlined in Chapter 3, data protection law could provide complementarity with non-discrimination law. Legal support for such an integrated approach can be found in the TFEU, where Article 8 TFEU, which guarantees gender mainstreaming and that ‘in all its activities, the Union shall aim to eliminate inequalities, and to promote equality, between men and women’. Article 10 TFEU also makes combating discrimination a transversal aim for the EU’s policy making. Thus, a transversal approach is not only necessary to adequately address algorithmic discrimination, but the European Union is also obliged to mainstream an equality perspective in its legal and policy-making activities.

4.4.3 Knowledge-based solutions

4.4.3.1 Awareness raising and research

To meet the challenges and make the most of the benefits related to algorithmic decision making, it has been emphasised that the human beings who are working with algorithms should be ‘given the knowledge and tools to comprehend and interact with AI systems to a satisfactory degree and, where possible, be enabled to reasonably self-assess or challenge the system’.⁷⁷⁹ This implies training and awareness-raising of the humans who plan the use of algorithms, are involved in their development and training, and make decisions based on the output of algorithms or in ‘human-machine’ teams.⁷⁸⁰ In particular, they should be made aware of the ethical and legal risks of algorithms and the various challenges and characteristics described above, and they should be provided with the relevant tools and mechanisms to deal with them.⁷⁸¹ In addition, the explainability of algorithmically supported decisions should be improved.⁷⁸² In this context, it has been suggested that such an explainability requirement should reflect existing legal standards of explainability such as the ‘right to good administration’ or ‘the duty to give reasons’, rather than mathematical explainability.⁷⁸³

More specifically, it is crucial to make sure that regulators, judges, economic players, the IT sector, and the society at large are sufficiently aware of the risks of algorithmic discrimination. Knowledge of the ways in which algorithms can reinforce existing structural patterns of discrimination and inequality is

779 Independent High Level Expert Group on Artificial Intelligence (AIHLEG) (2019), ‘Ethics Guidelines for Trustworthy AI’ (Brussels) 16.

780 See e.g. Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers’ Deputies, Appendix, para B.1.3. In addition, it could mean stakeholder participation, for example by ensuring that workers or minority groups are involved in the process of algorithmic decision-making; see Independent High Level Expert Group on Artificial Intelligence (AIHLEG) (2019), ‘Ethics Guidelines for Trustworthy AI’ (Brussels) 19.

781 Compare AIHLEG (2019), ‘Ethics Guidelines for Trustworthy AI’ (Brussels) 23, also suggesting that algorithms and AI should be developed by inclusive and diverse design teams, and proposing a ‘trustworthy AI assessment list’ to help stakeholders make decisions on algorithms and AI.

782 Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers’ Deputies, Appendix, para. B.4. See also eg Hamon, R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* (JRC Technical Report, EU Science Hub) (European Commission, Brussels) 24.

783 See Olsen, Henrik Palmer et al. (2021) ‘What’s in the Box? The Legal Requirement to Explain Computationally Aided Decision-Making in Public Administration’ in *Constitutional Challenges in the Algorithmic Society* (Oxford University Press, forthcoming).

crucial for bias mitigation, and only if there is such knowledge and awareness, tools and instruments can be used to address the risks related to algorithm-driven decision making. Since such risks of algorithmic discrimination can only be identified if patterns of social exclusion and disadvantage are known, this presupposes that such structural patterns of discrimination and inequality are researched, analysed, and widely exposed. Such knowledge will also enable subsequent steps (which are further addressed in section 4.4.4) such as ‘cleaning’ data to avoid bias, adapting data collection to non-discrimination requirements and monitoring algorithmic outcomes.

Such empirical (interdisciplinary) research, as well as research into which fields are most affected by algorithmic applications (as described in section 3.1), needs support and funding. EU and national agencies could have an important role to play in this regard. The scientific community itself may be called on to monitor and analyse the real effects of algorithms on discrimination and gender inequality. For example, it will have to face the crucial question of whether algorithms can create ‘new’ forms of discrimination and how these can best be prevented or redressed.

4.4.3.2 Databases, reporting tools and platforms

Knowledge about the (potential) risks and problems of discrimination related to algorithms can be further built up by means of mechanisms such as reporting tools, scoreboards, etc. The existence of such mechanisms could encourage watchdogs and whistleblowers to signal suspicions of algorithmic discrimination and draw public attention to problematic practices. Free and open databases could further support testing and assessment of potentially discriminatory algorithms.

In addition, section 4.4.2.4 has shown that various platforms already exist in which private and public actors cooperate to address problems of algorithmic discrimination. These platforms can also be used to share good practice and disseminate knowledge about effective ways to deal with algorithmic discrimination at the legal, societal and technological levels.

4.4.3.3 Training and education

Training and education are instrumental in tackling algorithmic discrimination. It is not only ICT specialists who should be trained to identify and prevent discrimination, but also all relevant professional communities including regulators, judges, equality bodies, etc. Educational institutions on all different levels (primary, secondary, tertiary, vocational) have a role to play in this. Good practices mapped out in section 4.3 for example include establishing university degrees or courses in ‘digital humanities’, digital social sciences, digital legal science, or access to on-the-job training and workshops on how to prevent algorithmic discrimination.

Similar to medical training, which includes an element of ethics training, ICT education programmes should train students to understand the risks of algorithmic discrimination and be able to devise and implement prevention strategies. In addition, knowledge-building and training should empower human agency in the human-machine relationship and allow human agents (regulators, judges, recruiters or anyone else tasked with algorithm-assisted decision making) to effectively respond to the challenge of automation bias. Knowledge of potential risks should enable active human control over algorithmic decisions. Consultation with users and multiple human agency would improve ‘human in the loop’ types of monitoring. Agency over algorithmic decisions by multiple trained humans would also minimise the risk of ‘rubber-stamping’.⁷⁸⁴

⁷⁸⁴ It has for example been suggested that human oversight in public administrative decisions made with the support of AI could take the form of an ‘administrative Turing test’, whereby decisions are validated by humans without them knowing whether a given output was generated by a machine or another human. See Olsen, Henrik Palmer et al. (2021) ‘What’s in the Box? The Legal Requirement to Explain Computationally Aided Decision-Making in Public Administration’.

As highlighted in section 4.3.3, diversity and participation in education and in employment are also key elements towards creating a more equality-compliant AI. Better inclusion of minorities in STEM curricula with a view to increasing the diversity of professional communities involved in developing and using algorithmic architectures is key to bias minimisation.⁷⁸⁵ In that regard, the whole spectrum of measures discussed in section 4.3.3 (ranging from quotas to targeted support) could help to close the current 'digital gender gap' and could be broadened to also improve the inclusion of other underrepresented groups.

4.4.4 Technology-based solutions

The six technological challenges discussed in Chapter 1 (the human factor and stereotyping and bias challenge; the data challenge; the correlation and proxy challenge; the transparency and explainability challenge; the scale and speed challenge, and the responsibility, accountability and liability challenge) demonstrate that algorithmic decision making is different from traditional human decision making. It is unavoidable that these specific technological characteristics and challenges, to a large degree, play a role in determining what is technically and realistically feasible in dealing with algorithmic discrimination. It is equally important, however, to realise that technology not only poses risks from a non-discrimination perspective, but also offers very important opportunities.⁷⁸⁶

4.4.4.1 Ex ante or preventive strategies

In particular, technologies are available that can help prevent some forms of discrimination, especially in the planning and development stages discussed in section 1.3 (*ex-ante* approaches).⁷⁸⁷ In these stages, first, equality and non-discrimination impact assessments could be undertaken and used to inform further decisions to be made on the technologies to be deployed.⁷⁸⁸ The European Institute for Gender Equality (EIGE) defines gender impact assessments as 'an ex ante evaluation, analysis or assessment of a law, policy or programme that makes it possible to identify, in a preventative way, the likelihood of a given decision having negative consequences for the state of equality between women and men'.⁷⁸⁹ Similar prevention strategies exist for other protected grounds in the form of equality impact assessments. Depending on the outcomes of such an impact assessment, it could be decided not to resort to an algorithm, or to take specific risks into account in developing it and in making choices as regards the data sets to be used to train or feed the algorithm, or about the use of the outputs of the algorithm in the eventual decision-making processes.

785 See e.g. <https://thriveglobal.com/stories/how-diversity-can-remove-cultural-bias-from-artificial-intelligence/>.

786 See in particular Kleinberg, J and others (2019), 'Discrimination in the Age of Algorithms', available at: <https://arxiv.org/abs/1902.03731>.

787 eg Kroll JA, Huey J and others (2017), 'Accountable Algorithms', 165 *University of Pennsylvania Law Review* 633; Žliobaitė, I (2017), 'Measuring discrimination in algorithmic decision making' 31 *Data Mining Knowledge Discovery* 1060, 1064.

788 In line with e.g. AIHLEG (2019), 'Ethics Guidelines for Trustworthy AI' (Brussels), where it is proposed to introduce 'fundamental rights impact assessments'; see also Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies, Appendix, para. B.5.2. See also Hamon R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* (JRC Technical Report, EU Science Hub) (European Commission, Brussels) 22.

789 See EIGE, 'Gender Impact Assessment' available at: <https://eige.europa.eu/gender-mainstreaming/toolkits/gender-impact-assessment/what-gender-impact-assessment>. See also e.g. Verloo, M and Roggeband, C (1996) 'Gender impact assessment: The development of a new instrument in the Netherlands' 14 *Impact Assessment* (1)3.

Secondly, to enhance fairness of algorithms in the development stage, use could be made of the work that is currently done to create ‘accountable algorithms’ and ‘equality by design’.⁷⁹⁰ Indeed, in its white paper on artificial intelligence, the European Commission has advised setting specific requirements in this regard in order to prevent outcomes entailing prohibited discrimination.⁷⁹¹ More specifically, several prevention methodologies in algorithmic design have been suggested to eliminate and mitigate bias and improve algorithmic design in relation to non-discrimination and gender equality standards.⁷⁹² In this context, various de-biasing strategies have been drafted by computer scientists.⁷⁹³ For example, ‘gender-tagging’ enables developers to make visible correlations found by an algorithm that may have a gendered connotation. This process allows humans to make decisions as to which correlations reflect an acceptable causal relationship or a justifiable differentiation, and which correlations are stereotyped or prejudiced and should be avoided.⁷⁹⁴ If the algorithm is trained accordingly, this could prevent gender biases or discrimination from tainting the output it generates. Such tagging can also be used in relation to other prohibited grounds of discrimination.⁷⁹⁵ Indeed, it may be necessary to permit the use of sensitive personal characteristics (e.g. ethnicity) directly in the process of developing and training an algorithm in order to make it possible to ‘sanitise’ the model.⁷⁹⁶ Thus, ‘fairness-aware’ algorithms may be created.⁷⁹⁷

A related preventive technological tool relates to the quality of the data used for developing and training an algorithm. If input data are biased and these biases go unnoticed, the algorithm might easily learn to recognise stereotyped patterns and generate a discriminatory output.⁷⁹⁸ Data augmentation – a strategy aimed at enhancing the diversity of data in a given dataset – or ‘cleaning’ strategies are available, however, that may help to de-bias the data in the process of selection and preparation of the relevant

-
- 790 See e.g. Kleinberg, J and others (2019), ‘Discrimination in the Age of Algorithms’, available at: <https://arxiv.org/abs/1902.03731>, 16; Kroll JA, Huey J and others (2017), ‘Accountable Algorithms’ 165 *University of Pennsylvania Law Review* 633; Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’ 55 *Common Market Law Review* 1143, 1171; Diakopoulos, N and Friedler, S (2016), ‘How to Hold Algorithms Accountable?’ *MIT Technology Review*, available at: <https://www.technologyreview.com/2016/11/17/155957/how-to-hold-algorithms-accountable/>; Mittelstadt, B (2016), ‘Auditing for transparency in content personalization systems’ 10 *International Journal of Communication* 4991; Sandvig, C and others (2014), ‘Auditing Algorithms: research methods for detecting discrimination on internet platforms’ Paper presented to *Data and Discrimination: Converting Critical Concerns into Productive Inquiry* (2014).
- 791 See e.g. European Commission (2020) *White Paper on Artificial Intelligence: A European approach to excellence and trust*, COM(2020) 65 final (Brussels 2020) 23. Similar recommendations have been made in the framework of the Council of Europe; see in particular Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers’ Deputies, Appendix, para B.5.
- 792 Hajian, S and others (2015), ‘Discrimination- and privacy-aware patterns’, 29 *Data mining and knowledge discovery* (6) 1733; Hajian, S and others (2014), ‘Generalization-based privacy preservation and discrimination prevention in data publishing and mining’, 28 *Data mining and knowledge discovery* (5-6) 1158; Hajian, S and Domingo i Ferrer, J (2013), ‘A methodology for direct and indirect discrimination prevention in data mining’, 25 *IEEE Transactions on knowledge and data engineering* (7) 1445.
- 793 Some of these opportunities might arise in particular in the development stage; for example, it may be easier to detect biases if a requirement exists for bias detection mechanisms to be built in the technology. For detailed examples of such techniques, see eg <https://aif360.mybluemix.net/>.
- 794 Compare Kleinberg, J and others (2019), ‘Discrimination in the Age of Algorithms’, available at: <https://arxiv.org/abs/1902.03731> 33; Kim PT (2017), ‘Data-Driven Discrimination at Work’ 58 *William and Mary Law Review* 857, 871.
- 795 See e.g. with many references, Khademi, A and others (2019), ‘Fairness in algorithmic decision making: an excursion through the lens of causality’ (27 March 2019), available at: <https://arxiv.org/pdf/1903.11719.pdf>; Kleinberg, J and others (2019), ‘Discrimination in the Age of Algorithms’, 5; Hacker P (2018), ‘Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law’, 1176 (also explaining some of the problems related to determining algorithmic ‘fairness’).
- 796 Žliobaitė, I and Custers, BHM (2016), ‘Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models’ 24 *Artificial Intelligence and Law* 183; cf also Hamon R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* (JRC Technical Report, EU Science Hub) (European Commission, Brussels) 18.
- 797 Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making: Opportunities and challenges* (Panel for the Future of Science and Technology (STOA) of the European Parliament) 46; compare also Hacker, P (2018), ‘Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law’, 1178.
- 798 Compare Katyal, SK (2019), ‘Private Accountability in the Age of Artificial Intelligence’ 66 *UCLA Law Review* 54, 68.

datasets.⁷⁹⁹ These strategies help to highlight and filter out any problematic biases and thus ‘clean up’ the training data.

4.4.4.2 Ex post strategies: screening and auditing algorithms

Importantly, *ex ante* strategies such as those discussed in the previous section may not always help reduce the risks of discrimination. Some machine-learning algorithms (in particular deep-learning algorithms) continue to adapt and develop once they have been trained, based on the real-life data that are fed into them. If these data reflect discrimination or biases – which they almost unavoidably do because of the prevalence of discrimination in our societies – these may creep into algorithms that are already in use, and affect their output.⁸⁰⁰ This process may be difficult to detect because of the opacity challenge, even for specialised supervisory bodies.⁸⁰¹ It therefore may be necessary to also introduce a number of *ex post* strategies, in particular in the form of auditing.⁸⁰² Research has been conducted into so-called ‘screening’ or ‘auditing algorithms’ which could detect the discriminatory potential of algorithmic patterns or output.⁸⁰³ Such tools could be used to systematically assess the discriminatory risks of algorithms before they are put to use or placed on the market, in a similar fashion to drug testing. In particular, these auditing algorithms can also help to pinpoint where the algorithm has gone wrong and may thus allow for targeted corrections. In addition, requirements could be introduced to keep a track record of the evolution of learning algorithms for testing and evidentiary purposes, for instance in case of legal proceedings, in particular for high-impact algorithms.

4.5 Conclusion: PROTECT – proposal for an integrated approach to algorithmic discrimination

The legal, knowledge-based and technical solutions and tools delineated in section 4.1 need to be deployed in complementary ways in order to allow them to help respond to the challenges of algorithmic discrimination. Below, we propose an integrated framework – PROTECT – to address the issue of algorithmic discrimination, taking account of the legal and practical challenges that have been described in this report and bringing together the different possible tools, instruments, solutions and good practice.

The integrated framework is based on the idea that a variety of strategies have to be put in place to address the problem of algorithmic discrimination. The prevention of algorithmic discrimination can be achieved through integrating various legal, knowledge-based and technological measures. These include the *diversification* of professional communities designing and training algorithms, the deployment of *equality by design strategies* offering guidance on the equality law framework to computer and data scientists, as well as the introduction of *equality and gender impact assessments* aimed at mainstreaming equality in algorithmic design.

There are two important prerequisites that have to be met in order to make such prevention strategies effective. First, *training* and *knowledge dissemination* about the inequality challenges faced by society are crucial. IT professionals should be educated in gender equality and non-discrimination law in the same

799 Hamon R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* 18; Katyal, SK (2019), ‘Private Accountability in the Age of Artificial Intelligence’, 68; Kroll JA, Huey J and others (2017), ‘Accountable Algorithms’ 165 *University of Pennsylvania Law Review* 633, 688; Crawford K and Whittaker M (2016), *The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term* available at: https://ainowinstitute.org/AI_Now_2016_Report.pdf, 6. On the process of preparing data for algorithmic use, see in general terms also Gillespie, T (2014), ‘The Relevance of Algorithms’ in Gillespie, T, Boczkowski, PJ and Foot, KA (eds), *Media technologies: Essays on communication, materiality, and society* (Cambridge, MIT Press) 167-194, 170.

800 See section 1.4.2.

801 See section 1.4.4.

802 AIHLEG (2019), ‘Ethics Guidelines for Trustworthy AI’ (Brussels) 19-20. Kroll, Huey and others also discuss several *ex post* approaches, such as review and oversight, see Kroll JA, Huey J and others (2017), ‘Accountable Algorithms’ 165 *University of Pennsylvania Law Review* 633.

803 See e.g. http://auditingalgorithms.science/?page_id=89#:~:text=Auditing%20Algorithms%3A%20Adding%20Accountability%20to,in%20diagnosing%20the%20unwanted%20consequences.

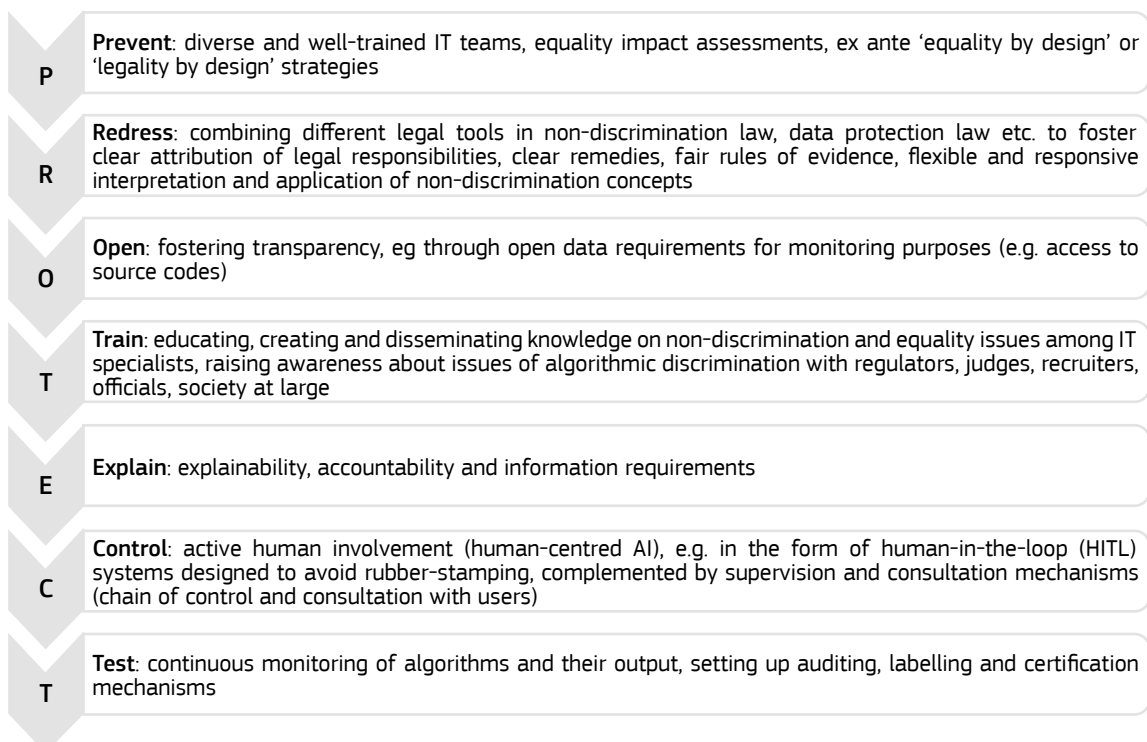
way as medical professionals receive ethical training. Conversely, equality law professionals (judges, regulators, equality bodies, etc.) as well as citizens and public and private users of algorithms should be informed of the discriminatory risks linked to the use of AI and of existing debiasing strategies. Second, prevention strategies are conditional on the *transparency* and *explainability* of algorithms. In the same vein, the availability of *open and clean data* for training and control purposes is key to prevention strategies.

As a continuous mechanism, constant monitoring is important to curb algorithmic discrimination. *Testing* mechanisms should be put in place to audit algorithms, in particular high-impact ones. *Certification* strategies could be put in place by tech companies in order to guarantee that the algorithms they design and sell are not discriminatory. Such monitoring strategies will have to be paired with efforts by tech companies to improve the *transparency*, *accountability* and *explainability* of algorithms. In line with the second dimension of the black box metaphor, the new horizons opened up by algorithmic technologies should be turned into opportunities to better detect and correct discrimination.

Human control also plays a vital role in this integrated approach to algorithmic discrimination. Public collective supervision as well as individual human supervision, combined with a clear allocation of liability and legal responsibility, will foster active human control over decisions relying on algorithmic recommendations or predictions. Such elements will discourage rubber-stamping and hopefully offset automation biases.

Finally, *legal redress* should be made available in the areas where it is lacking. The problem of algorithmic discrimination increases the weaknesses of EU equality law. Addressing algorithmic discrimination will mean reconsidering the gaps in the material scope of EU gender equality and non-discrimination. Adapting and revisiting some of the core concepts of the EU equality doctrine will also be necessary in order to accommodate the changing nature of discrimination. Legal redress will have to be transversal and integrate gender equality and non-discrimination law with other legal areas, not least data protection law. Cooperation between the relevant institutions on each side will also be vital in ensuring effective redress against algorithmic discrimination.

The PROTECT approach



General conclusions

This report has shown how the increasing use of algorithms in all areas of society poses challenges in terms of discrimination and raises important legal questions in Europe. First, this report has provided key definitions and information on the operation, creation, functions and types of algorithms, and it has outlined six major challenges linked to algorithmic discrimination: human bias, bias in data, proxy discrimination, the lack of transparency and explainability, the scale and speed of algorithmic decision making, and difficulties linked to accountability, responsibility and liability. The report has then turned to examining how these new challenges create frictions with the EU gender equality and non-discrimination legal framework. It has demonstrated how algorithmic discrimination risks falling into the cracks of EU gender equality and non-discrimination law because of existing gaps in the material scope, uncertainties and lack of flexibility in the personal scope, conceptual frictions, doctrinal mismatches, as well as procedural difficulties and enforcement challenges.

In addition to this, the report has reviewed the application of algorithmic techniques in national contexts and identified the specific problems that exist in relation to algorithmic decision making in the 31 European countries studied in this report. It has also mapped current public and scholarly discussions of the issue of algorithmic discrimination in these countries and identified existing national legal responses to algorithmic discrimination. Finally, the report has highlighted how algorithms can provide benefits and opportunities in the fight against gender inequality and discrimination by allowing us to better visualise, measure, detect and ultimately correct discriminatory biases if proper legal regulation and public policy is put in place. It has also mapped existing good practice at national and EU level in both public and private sectors.

The report has concluded by proposing a new integrated framework called PROTECT, which offers a set of potential legal, knowledge-based and technological measures and solutions to prevent, address and redress algorithmic discrimination.

All in all, this report demonstrates that addressing the question of algorithmic discrimination is key to building the 'ecosystem of trust' demanded by the EU white paper on artificial intelligence.⁸⁰⁴ This endeavour will require interdisciplinary efforts and a dialogue between law and policy makers and computer and data scientists. In addition, from a legal point of view, it can be concluded that many of the problems posed by algorithmic discrimination reinforce weaknesses and shortcomings that already exist in the legal framework. While these gaps and problems have already been the object of criticism for some time (e.g. the uneven material scope of EU discrimination law, the exceptions to gender equality in the field of goods and services, the comparator problem, the lack of recognition for intersectional discrimination, the exhaustive nature of the list of protected grounds, and uncertainties in the distinction between direct and indirect discrimination), the unprecedented scale and speed at which algorithmic discrimination might spread in the near future mean that addressing them is an urgent necessity.

If, as this report argues, the increasing use of algorithms in decision making in all areas of life risks systematically amplifying and magnifying existing structural inequalities, a piecemeal approach focused on the redress of individual instances of discrimination will not suffice. Neutrality towards algorithmic discrimination will only perpetuate the current discriminatory status quo. Instead, the approach that law and policy makers will need to adopt to effectively eradicate algorithmic discrimination is one that tackles the roots of discrimination with the aim of transforming the status quo. If 'bias in' means 'bias out', gender equality and non-discrimination laws and policies in Europe need to aim to actively eliminate such incoming biases. To this end, substantive and transformative equality approaches, which encompass instruments such as positive action measures and anti-stereotyping strategies, will prove essential. In the long run, it is only through debiasing society that we will be able to fully debias algorithms.

804 European Commission (2020), *White Paper on Artificial Intelligence: A European approach to excellence and trust* COM(2020) 65 final (Brussels 2020), 3.

Bibliography

Case law of the European Court of Justice

Judgment of 31 March 1981, *JP Jenkins v Kingsgate (Clothing Productions) Ltd* C-96/80 EU:C:1981:80

Judgment of 13 May 1986, *Bilka – Kaufhaus GmbH v Karin Weber von Hartz* C-170/84 EU:C:1986:204

Judgment of 17 October 1989, *Handels- og Kontorfunktionærernes Forbund i Danmark v Dansk Arbejdsgiverforening, acting on behalf of Danfoss* C-109/88 EU:C:1989:383

Judgment of 27 October 1993, *Dr. Pamela Mary Enderby v Frenchay Health Authority and Secretary of State for Health* C-127/92 EU:C:1993:859

Judgment of 30 April 1996, *P v S and Cornwall County Council* C-13/94 EU:C:1996:170,

Judgment of 13 January 2004, *Debra Allonby v Accrington & Rossendale College, Education Lecturing Services, trading as Protocol Professional and Secretary of State for Education and Employment* C-256/01 EU:C:2004:18

Judgment of 22 November 2005, *Werner Mangold v Rüdiger Helm* C-144/04 EU:C:2005:709

Judgment of 11 July 2006, *Sonia Chacón Navas v Eurest Colectividades SA* C-13/05 EU:C:2006:456

Judgment of 10 July 2008, *Centrum voor gelijkheid van kansen en voor racismebestrijding v Firma Feryn NV* C-54/07 EU:C:2008:397

Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law* C-303/06 EU:C:2008:415

Judgment of 11 November 2010, *Dita Danosa v LKB Līzings SIA* C-232/09 EU:C:2010:674

Judgment of 21 July 2011, *Patrick Kelly v National University of Ireland (University College, Dublin)* C-104/10 EU:C:2011:506

Judgment of 19 April 2012, *Galina Meister v Speech Design Carrier Systems GmbH* C-415/10 EU:C:2012:217

Judgment of 6 December 2012, *Johann Odar v Baxter Deutschland GmbH* C-152/11 EU:C:2012:772

Judgment of 25 April 2013, *Asociația Accept v Consiliul Național pentru Combaterea Discriminării* C-81/12 EU:C:2013:275

Judgment of 18 December 2014, *Fag og Arbejde (FOA) v Kommunernes Landsforening (KL)* C-354/13 EU:C:2014:2463

Judgment of 29 April 2015, *Geoffrey Léger v Ministre des Affaires sociales, de la Santé et des Droits des femmes and Etablissement français du sang* C-528/13 EU:C:2015:288

Judgment of 16 July 2015, *'CHEZ Razpredelenie Bulgaria' AD v Komisija za zashtita ot diskriminatsia* C-83/14 EU:C:2015:480

Judgment of 24 November 2016, *David L. Parris v. Trinity College Dublin and Others*, C-443/15, EU:C:2016:897; opinion of Advocate General Kokott, EU:C:2016:493

Judgment of 14 March 2017, *Samira Achbita and Centrum voor gelijkheid van kansen en voor racismebestrijding v G4S Secure Solutions NV* C-157/15 EU:C:2017:203

Judgment of 14 March 2017, *Asma Bougnaoui et Association de défense des droits de l'homme (ADDH) contre Micropole SA* C-188/15 EU:C:2017:204

Judgment of 6 April 2017, *Jyske Finans A/S v Ligebehandlingsnævnet, acting on behalf of Ismar Huskic* C-668/15 EU:C:2017:278

Judgment of 20 December 2017, *Asociación Profesional Elite Taxi v Uber Systems Spain, SL* C-434/15 EU:C:2017:981

Judgment of 17 April 2018, *Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung eV* C-414/16 EU:C:2018:257

Judgment of 15 November 2018, *Heiko Jonny Maniero v Studienstiftung des deutschen Volkes eV* C-457/17 EU:C:2018:912

Judgment of 11 September 2018, *IR v JQ* C-68/17 EU:C:2018:696

Judgment of 23 April 2020, *NH v Associazione Avvocatura per i diritti LGBTI — Rete Lenford* C-507/18 EU:C:2020:289

Case law of the European Court of Human Rights

Kiyutin v. Russia, Application no. 2700/10 (2011)

National case law

Finland

Case 216/2017, of 21.3.2018, https://www.yvtltk.fi/material/attachments/ytaltk/tapausselosteet/2SVkNzOWF/YVTltk-tapausseloste-_21.3.2018-luotto-moniperusteinen_syrjinta-S_L.pdf
Opinion of the Advertising Ethics Council MEN 35/2019 (Keskus-Kauppakamari (Finland Chamber of Commerce) 2019), available at: <https://kauppakamari.fi/statement-archive/men-35-2019-mainoksen-kohdentaminen-sosiaalisessa-mediassa/>

France

Administrative Guideline 24 April 2017 cancelled by the Conseil d'Etat (*Circulaire 24 avril 2017 annulé par le Conseil d'Etat*), 22 December 2017 Association SOS Education, Promotion et défense des étudiants et Droits des Lycéens Nos. 410561, 410641, 411913
Council of State No. 427916 12 June 2019, <https://www.conseil-etat.fr/ressources/decisions-contentieuses/dernieres-decisions-importantes/conseil-d-etat-12-juin-2019-universite-des-antilles>
Constitutional Council, Decision No. 2020-834 QPC 3 April 2020

Germany

Federal Labour Court, judgment of 22 July 2010, 8 AZR 1012/08, overruling the State Labour Court of Berlin and Brandenburg, judgment of 26 November 2008, 15 Sa 517/08
BGH, 14 May 2013, VI ZR 269/12 (*Google Auto-complete*)

Italy

TAR Lazio decisions No. 10964/2019, No. 9924/2018 and No. 6606/2019

Consiglio di Stato decisions: No. 881/2020; No. 30/2020; Nos. 8472, 8473, 8474/2019; No. 2936/2019, No. 2270/2019

Netherlands

ABRvS (Judicial Division of the Council of State) 17 May 2017, ECLI:NL:RVS:2017:1259 (Aerius I)

ABRvS 18 July 2018, ECLI:NL:RVS:2018:2454 (Aerius II)

HR (Supreme Court) 17 August 2018, ECLI:NL:HR:2018:1316 (WOZ)

Rechtbank (District Court) The Hague 5 February 2020, ECLI:NL:RBDHA:2020:865 (SyRI)

Poland

Supreme Administrative Court, judgment of 27 February 2014, case No. I OSK 2014/13

Judgment of 29 September 2016, reference number act II SA / Po 395/16

Warsaw District Administrative Court, case No. III SA / Wa 2057/18

Case K 53/16, Judgment of the Constitutional Tribunal of 6 June 2018, Journal of Law 2018/1149 (Wyrok Trybunału Konstytucyjnego z 6 czerwca 2018 r., sygn. K 53/16, Opublikowano: OTK-A 2018/38, Dz.U.2018/1149), available at: <https://sip.lex.pl/orzeczenia-i-pisma-urzedowe/orzeczenia-sadow/k-53-16-wyrok-trybunalu-konstytucyjnego-522591612>

Warsaw District Court, 5 September 2018, case No. II SAB / Wa 61/18

District Court of Poznań-Grunwald and Jeżyce, sitting in Poznań, partial judgment of 5 July 2018, case no. VP 695/16

Sweden

Equality Ombudsman, GRA 2017/80

Equality Ombudsman, TIL 2018/22

United Kingdom

Gibbs v. Westcroft [2014] 12 WLUK 110

GLS v. Brookes [2017] IRLR 780

R (Bridges) v. CCSWP and SSHD [2019] EWHC 2341 (Admin)

EU legislation and policy

Council Directive 79/7/EEC of 19 December 1978 on the progressive implementation of the principle of equal treatment for men and women in matters of social security [1978] OJ L 6/24

Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [1985] OJ L210/29

Council Directive 92/85/EEC of 19 October 1992 on the introduction of measures to encourage improvements in the safety and health at work of pregnant workers and workers who have recently given birth or are breastfeeding [1992] OJ L 348/1

Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin (Racial Equality Directive) [2000] OJ L 180/22

Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services (Gender Goods and Services Directive) [2004] OJ L 373/37

Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast) (Gender Recast Directive) [2006] OJ L 204/23

European Commission (2008) *Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation* COM(2008) 426 final OJ C303/8 (European Union, 2008)

Directive 2010/41/EU of the European Parliament and of the Council of 7 July 2010 on the application of the principle of equal treatment between men and women engaged in an activity in a self-employed capacity and repealing Council Directive 86/613/EEC [2010] OJ L 180/1

European Commission Recommendation 2014/124/EU of 7 March 2014 on strengthening the principle of equal pay between men and women through transparency, OJ L 69, 8.3.2014

Regulation 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1

European Commission (2018) Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 'Artificial Intelligence for Europe', COM(2018) 237 final (Brussels 2018)

European Commission (2018) Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions 'Coordinated Plan on Artificial Intelligence', COM(2018) 795 final (Brussels 2018)

Council of the European Union (2019) 'Conclusions on the coordinated plan on artificial intelligence' (11 February 2019) (6177/19) available at: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/en/pdf>

Act of 21 February 2019 amending certain acts in connection with ensuring the application of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on free flow of such data and repealing Directive 95/46 / EC (General Data Protection Regulation) (Journal of Laws, item 730)

Directive 2019/1158/EU of the European Parliament and of the Council of 20 June 2019 on work-life balance for parents and carers and repealing Council Directive 2010/18/EU (Work-Life Balance Directive) [2019] OJ L 188/79

European Commission (2020) *White Paper on Artificial Intelligence: A European approach to excellence and trust*, COM(2020) 65 final (Brussels 2020)

European Commission (2020) *A Union of Equality: Gender Equality Strategy 2020-2025*, COM/2020/152 final

Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies, Appendix

National legislation and policy documents

Belgium

Collège des procureurs généraux / College van procureurs-generaal (College of Public Prosecutors), Circular COL. 15/2018 adopted on 20 December 2018 (confidential) available at: <https://www.om-mp.be/fr/savoir-plus/circulaires>

Bulgaria

Law on Electronic communications (Закон за електронните съобщения/-S.G.) No. 41/ 2007, last amended on 5 June 2020- <https://www.lex.bg/laws/ldoc/2135553187>

National Strategy for integration of Roma 2012-2020, see https://ec.europa.eu/info/sites/info/files/roma_bulgaria_strategy_en.pdf

Croatia

Criminal Code (*Kazneni zakon*), Official Gazette *Narodne novine* Nos. 28/2013, 33/2015, 82/2015, 82/2016, 67/2018 and 126/2019

Czechia

Working group on Human Rights and New Technologies (2019), *Regulatory Framework for Artificial Intelligence in the European Union* (Position paper on the EU White Paper), available at: http://observatory.ilaw.cas.cz/wp-content/uploads/2020/02/CZ%20nonpaper_%20EU%20regulatory%20framework%20for%20AI.pdf

Denmark

European Affairs Committee of the Danish Parliament (2020), available at: www.ft.dk/samling/20191/almindel/ERU/bilag/210/2171335/index.htm

France

Ministerial order by Ministry of Finance of 28 August 2017 creating an automated anti-fraud system entitled 'fraud targeting and request enhancement' (*Arrêté du 28 août 2017 modifiant l'arrêté du 21 février 2014 portant création par la direction générale des finances publiques d'un traitement automatisé de lutte contre la fraude dénommé « ciblage de fraude et valorisation des requêtes »*)

Ministry of Innovation and Digital Issues and Ministry of Education (2017), *Rapport de synthèse, France Intelligence Artificielle* (Summary Report, France Artificial Intelligence) available at: www.cnil.fr/sites/default/files/atoms/files/rapport_de_synthese_france_ia.pdf

Bill on bioethics, No. 686 of the General Assembly on August 3 2020 (see article 11) was sent to the Senate (second reading) now pending in front of the Mixed Commission in charge of drafting a version which integrates the amendments of the Senate and General Assembly, see: www.senat.fr/leg/pjl19-686.html; last version of the bill available at: www.senat.fr/leg/pjl19-686.html

Bill to fight against hate speech on internet, platforms and search engines in 24 hours adopted by the National Assembly 13 May 2020 but cancelled in most part by the Constitutional Council as an infringement of freedom of expression considering the difficulty to monitor all content closely, Conseil Constitutionnel, Decision No. 2020-801 DC of 18 June 2020 available at: www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm

Bill aimed to fight against hate contents on internet (*Loi du 24 juin 2020 visant à lutter contre les contenus haineux sur internet*) adopted by the National Assembly 13 May 2020, but cancelled in part by the Constitutional Council as an infringement of freedom of expression considering the difficulty to

monitor all content closely, Constitutional Council, Decision No. 2020-801 DC of 18 June 2020, available at: www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm

Germany

General Equal Treatment Act

Federal Government (2020), 'Strategie Künstliche Intelligenz der Bundesregierung' (National Strategy on Artificial Intelligence), available at: www.ki-strategie-deutschland.de/home.html

Greece

Act 3896/2010, OJ A 207/18.12.2010, transposing Directive 2006/54/EC

Act 3769/2009, OJ A 105/01.07.2009, transposing Directive 2004/113/EC

Act 4097/2012, OJ A 235/03.12.2012, transposing Directive 2010/41/EU

Act 4443/2016, OJ A 232/09.12.2016, transposing Directives 2000/43/EC and 2000/78/EC

Act 4604/2019, OJ A 50/26.03.2019

Italy

Agenzia per L'Italia Digitale (AGID) (Agency for Digital Italy) (2018), *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino* (White paper on AI at the service of the citizen) (Version 1.0) available at: ia.italia.it/assets/librobianco.pdf

Digital Administration Code (*Codice dell'Amministrazione Digitale*), Legislative Decree no. 82 of 7 March 2005 (*Decreto Legislativo 7 marzo 2005, n. 82*) available at: <https://docs.italia.it/italia/piano-triennale-ict/codice-amministrazione-digitale-docs/it/v2017-12-13/index.html>

Ministry of Economic Development (2019), *Strategia Nazionale per l'Intelligenza Artificiale* (National Strategy for Artificial Intelligence) available at: www.mise.gov.it/images/stories/documenti/Strategia-Nazionale-Intelligenza-Artificiale-Bozza-Consultazione.pdf

Latvia

Informative report of the Ministry of Environmental Protection and Regional Development Ministry on development of artificial intellect (*VARAM Informatīvais ziņojums Par mākslīgā intelekta risinājumu attīstību*) 51, available at: www.varam.gov.lv/lv/jauns-informativais-zinojums-par-maksliga-intelekta-risinajumu-attistibu

Lithuania

Lithuanian Artificial Intelligence Strategy (March 2019) available at: [http://eimin.lrv.lt/uploads/eimin/documents/files/DI_strategija_ENG\(1\).pdf](http://eimin.lrv.lt/uploads/eimin/documents/files/DI_strategija_ENG(1).pdf)

Malta

ACT No. XXXIII of 2018, 20 July 2018 <https://mdia.gov.mt/wp-content/uploads/2018/10/ITAS.pdf>

Ethical AI Framework (October 2019), available at: https://malta.ai/wp-content/uploads/2019/10/Malta_Towards_Ethical_and_Trustworthy_AI_vFINAL.pdf

Malta: The Ultimate AI Launchpad. A Strategy and Vision for Artificial Intelligence in Malta 2030 (Parliamentary Secretary for Financial Services, Digital Economy and Innovation, October 2019), available at: https://malta.ai/wp-content/uploads/2019/11/Malta_The_Ultimate_AI_Launchpad_vFinal.pdf

The Netherlands

Decision of 1 September pertaining to the amendment of the SUWI Decree in relation to rules for combating fraud by means of data exchange and the effective use of data known to the government using SyRI (*Besluit van 1 september 2014 tot wijziging van het Besluit SUWI in verband met regels*

voor fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekend zijnde gegevens met inzet van SyRI), *Staatsblad* (Official Journal) 2014, 320) available at: <https://zoek.officielebekendmakingen.nl/stb-2014-320.html>

Advisory Committee on the Implementation of Benefits (2020), *Omzien in verwondering. Eindadvies* (Looking back in amazement. Final advice) (part 2, The Hague, 12 March 2020), annex No. 926526 to *Kamerstukken II* 2019/20, 31066, No. 608 available at: www.rijksoverheid.nl/documenten/kamerstukken/2020/03/12/omzien-in-verwondering-eindadvies-adviescommissie-uitvoering-toeslagen

Letter of 23 April 2020 to the President of the House of Representatives by the State Secretary for Social Affairs and Employment, Tamara van Ark, on a court judgment regarding SyRI (*Kamerbrief van 23 april 2020 naar aanleiding van vonnis rechter inzake SyRI*), *Kamerstukken II* 2019/20, 17050, No 593 available at: www.rijksoverheid.nl/documenten/kamerstukken/2020/04/23/kamerbrief-naar-aanleiding-van-vonnis-rechter-inzake-syri

The Netherlands Government's policy brief of 8 October 2019 on offering guarantees against the risks of data-analyses by public bodies, available at: www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2019Z19084&did=2019D39751

Government of the Netherlands (2019), *Strategisch Actieplan Artificiële Intelligentie (SAPAI)* (Strategic Action Plan Artificial Intelligence (SAPAI)) (October 2019) available at: www.rijksoverheid.nl/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie

Minister of Economic Affairs and Climate Policy (2020), 'Kabinetsappreciatie witboek over Kunstmatige intelligentie' (Cabinet appreciation white paper on artificial intelligence) (13 May 2020), annex No. 932356 to *Kamerstukken II* 2019/20, 26643, No. 680

Norway

Minister of Local Government and Modernisation (2020), *National Strategy for Artificial Intelligence* (14 January 2020) available at: www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/

Minister of Local Government and Modernisation (2019), 'Erklæring om ansvarlig bruk av kunstig intelligens i arbeidslivet' (Declaration on the responsible use of artificial intelligence in working life) – A declaration signed by various companies and organisations in the private and public sector in Norway available at: www.regjeringen.no/contentassets/0e36c85fcfe143a5b626c53cf292cb3b/negotia_telenor_ansvarlig-bruk-av-kunstig-intelligens-i-arbeidslivet.pdf

Poland

Civil Code, Act of 23 April 1964, consolidated text, JoL of 2018, item 1025, as amended

Ministry of Digital Affairs (2019), *Polityka Rozwoju Sztucznej. Inteligencji w Polsce na lata 2019 – 2027* (Policy for the Development of Artificial Intelligence in Poland for 2019–2027)

Slovakia

Act No. 300/2005 Coll. Criminal Code

Action plan for the digital transformation of Slovakia for 2019–2022, adopted by a Government Resolution No. 337/2019 of 3 July 2019, available at: www.vicpremier.gov.sk/wp-content/uploads/2019/10/AP-DT-English-Version-FINAL.pdf

Bieliková, M and others (2019), *Analýza a návrh možností výskumu, vývoja a aplikácie umelej inteligencie na Slovensku* (Analysis and design of possibilities of research, development and application of artificial

intelligence in Slovakia) (11 December 2019, Slovak Technical University of Bratislava and Office of the Vice-President of the Government of the Slovak Republic for Investment and Informatisation) available in Slovak at: www.mirri.gov.sk/wp-content/uploads/2020/03/Dielo2-Manual.pdf

Spain

State Secretary for Digital Advancement (2020), *Libro Blanco de las mujeres en el ámbito tecnológico* (White paper on women in technology) available at: www.mineco.gob.es/stfls/mineco/ministerio/ficheros/libreria/LibroBlancoFINAL.pdf

Sweden

Discrimination Act (2008:567)

United Kingdom

Data Ethics Framework (published 13 June 2018; last updated 16 September 2020), www.gov.uk/government/publications/data-ethics-framework

Housing Benefit and Council Tax Benefit Circular, HB/CTB S11/2011 available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633018/s11-2011.pdf

House of Lords Select Committee on Artificial Intelligence (2018) *AI in the UK: Ready, Willing and Able?* (HL Paper 100, 16 April 2018) available at: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>

House of Commons Science and Technology Committee (2018), *Algorithms in Decision Making (Fourth Report of Sessions 2017-2019)* (HC 351, 15 May 2018) available at: <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/35102.htm>

AI Sector Deal (Policy Paper) (updated 21 May 2019), <https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal>

Literature, reports and other sources

AlGHLEG (Independent High-Level Expert Group on Artificial Intelligence) (2019) 'Ethics Guidelines for Trustworthy AI' (Brussels)

Aishow (2017), 'El Gobierno crea un Grupo de Sabios para elaborar un libro blanco sobre Inteligencia Artificial y Big Data' (The Government sets up a group of experts to produce a white book on artificial intelligence and big data) (15 November 2017) available at: <https://the-aishow.com/news/gobierno-crea-grupo-sabios-elaborar-libro-blanco-inteligencia-artificial-big-data/>

Ali, M and others (2019), 'Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes' arXiv preprint, arXiv:190402095

Allen, R (2020), 'Artificial intelligence, machine learning, algorithms and discrimination law: the new frontier' (Discrimination Law Conference, Michael Rubenstein Conferences, London, 31 January 2020) available at: https://482pe539799u3ynseg2hl1r3-wpengine.netdna-ssl.com/wp-content/uploads/2020/02/Discrimination-Law-in-2020.FINAL_-1.pdf

Allen, R and Masters, D (2020), 'Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making' 20(4) *ERA Forum* 585

Allen, R and Masters, D (2020), *Regulating for an equal AI: a new role for equality bodies. meeting the new challenges to equality and non-discrimination from increased digitisation and the use of artificial intelligence* (Equinet, Brussels) available at: https://equineteurope.org/wp-content/uploads/2020/06/ai_report_digital.pdf

Allhutter, D and others (2020), 'Algorithmic profiling of job seekers in Austria: how austerity politics are made effective', *Frontiers in Big Data*, available at: <https://www.frontiersin.org/articles/10.3389/fdata.2020.00005/full>

Amrani-Mekki, S (2018), 'Justice predictive et accès au juge' (Predictive justice and access to the judge) in Amrani-Mekki, S and others, *La justice prédictive. Actes du colloque du 12 février 2018 organisé par l'Ordre des avocats au Conseil d'Etat et à la Cour de cassation à l'occasion de son bicentenaire, en partenariat avec l'Université Paris-Dauphine PSL* (Predictive justice. Proceedings of the symposium of 12 February 2018 organised by the Bar Association at the Council of State and the Court of Cassation on the occasion of its bicentennial, in partnership with the University Paris-Dauphine PSL) (Daloz)

Apostolidis, C (2019), 'A legal approach to algorithm-driven process and artificial intelligence' (LLM thesis, International Hellenic University) available at: https://repository.ihu.edu.gr/xmlui/bitstream/handle/11544/29440/c.apostolidis_llm_16.04.2019.pdf?sequence=1

Arbeitsmarktservice Österreich, Nr.: GEZ: BGS/VOR/003/2019 of 25 April 2019, response to inquiry by the National Equality Ombuds

Arnardóttir, OM (2003), *Equality and non-discrimination under the European Convention on Human Rights* (Martinus Nijhoff Publishers)

Astobiza, AM (2017), 'Ética algorítmica: Implicaciones éticas de una sociedad cada vez más gobernada por algoritmos' (Algorithmic ethics: ethical implications of a society increasingly governed by algorithms) *24 Dilemata* 185

Autoriteit Persoonsgegevens (Tax and Customs Authority) (2020), *Belastingdienst/Toeslagen: De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag* (The processing of the nationality of recipients of child care benefits) (Onderzoeksrapport z2018-22445) available at: www.autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf

Aylward, C (2010), 'Intersectionality: Crossing the Theoretical and Praxis Divide' 1 *Journal of Critical Race Theory* 1

Bagger, T (2019), 'AUFF starting grant of 2,200,000 kr. for Vincenzo Pietrogiovanni from Department of Law', *Juridisk Institut*, (19 December 2019) available at: <https://law.medarbejdere.au.dk/aktuelt/nyheder/nyhed/artikel/auff-starting-grant-of-2200000-kr-to-vincenzo-pietrogiovanni-from-department-of-law/>

Balkin, JM (2017), '2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data' *Ohio State Law Journal* 1217

Baratta, L (2017), 'Intervista. La proposta della Cgil: l'algoritmo di Deliveroo e Foodora crea discriminazioni, contrattiamolo' (Interview. The proposal of Cgil: the algorithm of Deliveroo and Foodora creates discrimination, let's contract it) *Linkiesta*, (14 December 2017) <https://www.linkiesta.it/2017/12/la-proposta-della-cgil-lalgoritmo-di-deliveroo-e-foodora-crea-discrimi/>

Barocas, S and Selbst, A (2016), 'Big Data's Disparate Impact' 104 *California Law Review* 761

de Barros Fritz, R and von Lewinsky, K (2018), 'Arbeitgeberhaftung nach dem AGG infolge des Einsatzes von Algorithmen bei Personalentscheidungen' (Employer liability due to the use of algorithms in personnel decisions) *Neue Zeitschrift für Arbeitsrecht* 620

Barzilay, AR and Ben-David, A (2017), 'Platform Inequality: Gender in the Gig-Economy' 47 *Seton Hall Law Review* 393

Befring, AK (2019), *Persontilpasset medisin. Rettslige perspektiver* (Personalised medicine. Legal perspectives) (Gyldendal) available at: www.gyldendal.no/Faglitteratur/Jus/Juridiske-fag/Persontilpasset-medisin

BEIS (2018), 'Government response to House of Lords Artificial Intelligence Select Committee's Report on AI in the UK: Ready, Willing and Able?' (CM 9645, June 2018) available at: <https://www.gov.uk/government/publications/ai-in-the-uk-ready-willing-and-able-government-response-to-the-select-committee-report>

Bellamy, RKE and others (2018), 'AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias' arXiv:181001943 [csAI]

Belmonte, E (2019), 'La aplicación del bono social del gobierno niega la ayuda a personas que tienen derecho a ella' (The implementation of the Government's social bonus refuses support to individuals who are entitled to it), *Civio* (16 May 2019), <https://civio.es/tu-derecho-a-saber/2019/05/16/la-aplicacion-del-bono-social-del-gobierno-niega-la-ayuda-a-personas-que-tienen-derecho-a-ella/>

Benítez Eyzaguirre, L (2019), 'Ética y transparencia para la detección de sesgos algorítmicos de género' (Ethics and transparency for detection of gender bias in algorithms) 25(3) *Estudios sobre el mensaje periodístico* 1307

Berghahn, S, Klapp, M and Tischbirek, A (2016), *Evaluation des Allgemeinen Gleichbehandlungsgesetzes* (Evaluation of the General Equal Treatment Act) (ADS, Berlin), http://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/AGG/agg_evaluation.pdf?__blob=publicationFile&v=15

Bertail, P and others (2019), *Algorithmes : biais, discrimination et équité* (Algorithms: Bias, Discrimination and Equity) (Telecom ParisTech and Fondation Abeona, February 2019) available at: www.telecom-paris.fr/wp-content/uploads/2019/02/Algorithmes-Biais-discrimination-equite.pdf

Besson, S (2008), 'Gender Discrimination under EU and ECHR Law: never shall the twain meet?' 8 *Human Rights Law Review* (4) 647

Beuth, P (2017), 'Maas schlägt digitales Antidiskriminierungsgesetz vor' (Maas proposes a digital non-discrimination law) *Zeit Online* (3 July 2017) available at: www.zeit.de/digital/internet/2017-07/heiko-maas-algorithmen-regulierung-antidiskriminierungsgesetz

BiSmart (2020) 'Motivētākās sievietes aicina pieteikties bezmaksas apmācībām IT jomā' (The most motivated women are invited to apply for free of charge studies in IT sector) (25 April 2020), available at: <https://bismart.lv/blogs/vieda-valsts/sievietes-acinatas-pieteikties-bezmaksas-programmesanas-apmacibam-1420>

de Blasio, B and Menin, J (2015), 'From cradle to cane: the cost of being a female consumer. A study of gender pricing in New York City' The New York Department of Consumer Affairs

Bolukbasi, T and others (2016), 'Man is to computer programmer as woman is to homemaker? Debiasing word embeddings' *Advances in Neural Information Processing Systems* 4356

Bourcier, D and De Filippi, P (2018), 'Transparence des algorithmes face à l'open data : quel statut pour les données d'apprentissage?' (Transparency of algorithms in light of open data: what is the status of learning data?) 167 *Revue française d'administraton publique* (3)

Breemen, VE and Wouters, AHH (2020), 'Casestudy Zelfrijdende auto's' (Case study automated cars), in Kulk, S, van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek* (Legal aspects of algorithms that make decisions. An exploratory study) (The Hague, WODC)

Bribosia, E, Rorive, I and Hislair, J (2017), 'Article 21: Non Discrimination' in Picod, F, Van Drooghenbroeck, S and Rizcallah, C (eds), *Charte des droits fondamentaux de l'Union européenne: Commentaire article par article* (Charter of Fundamental Rights of the European Union: Article-by-article commentary) (Brussels, Bruylant)

Buolamwini, J and Gebru, T (2018), *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification* (Proceedings of Machine Learning Research)

van den Brink, M and Dunne, P (2018), *Trans and intersex equality rights in Europe – a comparative analysis* (Universiteit Utrecht)

Burrell, J (2016), 'How the machine "thinks": Understanding opacity in machine learning algorithms' 3 *Big Data & Society*

Burri, S and Heeger-Hertter, S (2018), 'Discriminatie in de platformeconomie juridisch bestrijden: geen eenvoudige zaak' (Combating discrimination in the platform economy with legal means: No simple matter) *Ars Aequi* 1000

Burri, S and McColgan, A (2008), *Sex-segregated Services* (European Commission)

Burri, S and McColgan, A (2009), *Sex Discrimination in the Access to and Supply of Goods and Services and the Transposition of Directive 2004/113/EC* (European Commission)

Busch, C (2018), *Algorithmic Accountability* (Expert report for the ABIDA – Assessing Big Data project funded by the German Federal Ministry for Education and Research) available at: www.abida.de/sites/default/files/ABIDA%20Gutachten%20Algorithmic%20Accountability.pdf

Busstra, M.J. (2011) *The implications of the Racial Equality Directive for minority protection within the European Union*. (Eleven International Pub)

Cadiet, L (2017), *L'Open data des décisions de justice. Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice* (Open data of court decisions. Mission of study and prefiguration on the opening to the public of court decisions) (Rapport à Madame la garde des Sceaux, ministre de la Justice, November 2017) available at: www.justice.gouv.fr/publication/open_data_rapport.pdf

Caracciolo di Torella, E and McLellan, B (2018), *Gender Equality and the Collaborative Economy* (Publications Office of the European Union)

Carballar, O (2020), 'Los algoritmos pueden intensificar la desigualdad' (Algorithms may render inequality stronger) *La Marea* (9 February 2020) available at: www.lamarea.com/2020/02/09/big-data-privacidad-desigualdad/

Castelluccia, C and Le Métayer, D (2019), *Understanding algorithmic decision-making: Opportunities and challenges* (Report for the Panel for the Future of Science and Technology (STOA) of the European Parliament)

Castelvecchi, D (2016), 'The Black Box of AI' *Nature* (538) 21

Del Castillo, C (2019), 'Carlos Preciado: "Los algoritmos a veces dan una falsa impresión de objetividad que aplasta los derechos de las personas"' (Carlos Preciado: 'Sometimes algorithms give a false impression of objectivity that quashes people's rights') *El Diario* (9 July 2019) available at: www.eldiario.es/tecnologia/algoritmo-maquinas-impresion-objetividad-derechos_0_918658323.html

Catanzaro, M (2020), 'In Spain, the VioGén algorithm attempts to forecast gender violence', *Algorithm Watch* (27 April 2020) available at: <https://algorithmwatch.org/en/story/viogen-algorithm-gender-violence/>

Cavallaro, MC and Smorto, G (2019), 'Decisione pubblica e responsabilità dell'amministrazione nella società dell'algoritmo' (Public decision and responsibility of the administration in the algorithm society), *Federalismi.it* (4 September 2019) available at: http://www.federalismi.it/nv14/articolo-documento.cfm?Arid=40182&content=Decisione%2Bpubblica%2Be%2Bresponsabilità%2Bdell'amministrazione%2Bnella%2Bsocietà%2Bdell'algoritmo&content_author=%3Cb%3EMaria%2BCristina%2BCavallaro%2Be%2BGuido%2BSmorto%3C%2Fb%3E

Chancerelle, L (2019), 'La lutte contre les discriminations en Europe à l'ère de l'intelligence artificielle et du big data' 381 *Journal du Droit des Jeunes* 25

Chellet, P (2020), 'Prix personnalisé sur internet : la rencontre de la discrimination, de la concurrence et des données personnelles', in Mercat-Bruns, M (ed), *Nouveaux modes de detection et de prevention de la discrimination et accès au droit: action de groupe et discrimination systémique, algorithmes et préjugés, réseaux sociaux et harcèlement* (New ways of detecting and preventing discrimination and access to the law. Class action and systemic discrimination. Algorithms and prejudices. Social networks and harassment) (1st edn, Société de législation comparée ed, Collection Trans Europe Experts, Vol 14) available at: www.lgdj.fr/nouveaux-modes-de-detection-et-de-prevention-de-la-discrimination-et-acces-au-droit-9782365170925.html

Chiusi, F and others (2020) *Automating Society Report 2020* (Algorithm Watch and Bertelsmann Stiftung), 34 available at: <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/10/Automating-Society-Report-2020.pdf>

Citron, DK and Pasquale, FA (2014), 'The Scored Society: Due Process for Automated Predictions' *Washington Law Review* 1

Le Clainche, J and Le Métayer, D (2012), 'Données personnelles, vie privée et non-discrimination: des protections complémentaires, une convergence nécessaire' (Personal data, privacy and non-discrimination: complementary protections, a necessary convergence), *Revue Lamy Droit immatériel* 90

Cluzel, L (2020), 'Décision publique algorithmique et discrimination' in Mercat-Bruns, M (ed), *Nouveaux modes de detection et de prevention de la discrimination et accès au droit : action de groupe et discrimination systémique, algorithmes et préjugés, réseaux sociaux et harcèlement* (New ways of detecting and preventing discrimination and access to the law. Class action and systemic discrimination. Algorithms and prejudices. Social networks and harassment) (1st edn, Société de législation comparée, Collection Trans Europe Experts, Vol 14) available at: www.lgdj.fr/nouveaux-modes-de-detection-et-de-prevention-de-la-discrimination-et-acces-au-droit-9782365170925.html

Collins, H and Khaitan, T (2018), 'Indirect Discrimination Law: Controversies and Critical Questions' in Collins, H and Khaitan, T (eds), *Foundations of Indirect Discrimination Law* (1 edn, Hart Publishing)

Corbett-Davies, S, Goel, S, and González-Bailón, S (2017), 'Even Imperfect Algorithms Can Improve the Criminal Justice System. A way to combat the capricious and biased nature of human decisions', *New York Times* 20 December 2017, available at: <https://nyti.ms/2Deshvp>

Cormen, TH and others (2009), *Introduction to Algorithms* (Cambridge, MIT Press)

Cotino Hueso, L (2017), 'Big Data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales' (Big Data and Artificial Intelligence. An Approach from a Legal Point of View about Fundamental Rights) 24 *Dilemata* 131

Crawford, K (2013), 'The Hidden Biases in Big Data' *Harvard Business Review* 2013, available at: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>

Crawford, K and Whittaker, M (2016), *The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term* available at: https://ainowinstitute.org/AI_Now_2016_Report.pdf

CSSF – Commission de Surveillance du Secteur Financier (2018), *White Paper on Artificial Intelligence. Opportunities, risks and recommendations* available at: www.cssf.lu/en/document/white-paper-artificial-intelligence-opportunities-risks-and-recommendations-for-the-financial-sector/

Czák, A (2019), 'Das Problem mit dem AMS-Algorithmus' (The Problem with the AMS-Algorithm) *Epicenter. works*, (17 October 2019) available at: <https://epicenter.works/content/das-problem-mit-dem-ams-algorithmus>

Criado Perez, C (2019) *Invisible women: Exposing data bias in a world designed for men* (Random House)

Gulbrandsen, C (2017), 'Rekrutterer ved hjelp av robot – ansetter flere kvinner' (Recruits using a robot – hires more women), *Dagens næringsliv* (7 December 2017) available at: <https://www.dn.no/arbeidsliv/robotteknologi/utdannelse/evry-evry/rekrutterer-ved-hjelp-av-robot-ansetter-flere-kvinner/2-1-227334>

Danish Expert Group on Data Ethics (2018), *Data for the Benefit of the People. Recommendations from the Danish Expert Group on Data Ethics* (November 2018), available at: <https://dataetiskraad.dk/sites/default/files/2020-02/Recommendations%20from%20the%20Danish%20Expert%20Group%20on%20Data%20Ethics.pdf>

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/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

Data Ethics Commission (2019), *Expertise*, available at: <https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenethikkommission/arbeitsergebnisse-der-dek/arbeitsergebnisse-der-dek-node.html>

Data Ombudsman's Office (2019), *Automaattinen päätöksenteko ja profilointi* (Automated decision making and profiling), available at: <https://tietosuoja.fi/automaattinen-paatoksenteko-profilointi>

Defender of Rights (2015), *Guide: Recruter avec des outils numériques sans discriminer* (Guide: Recruiting with digital tools without discrimination) available at: <https://www.defenseurdesdroits.fr/fr/outils/recruter-avec-des-outils-numeriques-sans-discriminer>

Defender of Rights and CNIL (National Data Protection Authority of France) (2020), *Algorithmes: prévenir l'automatisation des discriminations* (Algorithms: preventing automated discrimination), joint report, 7, available at: www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-algos-en-num-16.07.20.pdf

Defender of Rights (2017), 'Lutte contre la fraude aux prestations sociales : à quel prix pour les droits des usagers ?', (Fighting social benefits fraud: what price for users' rights?) 20, available at: https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=16746

Delépine, J (2019), 'Algoritmos y discriminación sexista' (Algorithms and sexist discrimination) 69, *Alternativas económicas* 23

Department of Business, Enterprise and Innovation (2019), 'Four in ten Irish companies currently use Artificial Intelligence (AI)' (8 April 2019) available at: <https://dbei.gov.ie/en/News-And-Events/Department-News/2019/April/09042019.html>

Depraetere, M, sp.a (2019), *Algorithmes dans le secteur des assurances* (Algorithms in the insurance sector), 6 November 2019, Question et réponse écrite No. 0122, législature: 55

Devillers, L (2017), *Des robots et des hommes* (On robots and men) (Plon)

Diakopoulos, N and Friedler, S (2016), 'How to Hold Algorithms Accountable?' *MIT Technology Review*, available at: <http://www.technologyreview.com/2016/11/17/155957/how-to-hold-algorithms-accountable/>

Van Dijck, J, Poell, T and De Waal, M (2018), *The Platform Society. Public Values in a Connective World* (Oxford, Oxford University Press)

Diskriminerings ombudsmannen (DO) (2019), 'Om behovet av effektiva och avskräckande sanktioner vid vissa överträdelser av diskrimineringslagen (2008:567) med mera' (On the need for effective and deterrent sanctions concerning certain violations of the Discrimination Act (2008:567) etc.) (16 December 2019) available at: www.do.se/om-do/vad-gor-do/skrivelser-till-regeringen/om-behovet-av-effektiva-och-avskrackande-sanktioner-vid-vissa-overtradelser-av-diskrimineringslagen-2008567-med-mera/

Diskriminerings ombudsmannen (DO) (2019), 'Vi behöver mer effektiva och avskräckande sanktioner' (We need more effective and deterrent sanctions) (Press release, 18 December 2019) available at: www.do.se/om-do/pressrum/aktuellt/aktuellt-under-2019/vi-behovet-mer-effektiva-och-avskrackande-sanktioner/

Dondero, B (2017), 'Justice predictive: la fin de l'aléa judiciaire?' *Recueil Dalloz* 532

Doshi-Velez, F, Korz, M and others (2017), 'Accountability of AI under the law: the role of explanation' (Berkman Klein Center for Internet & Society working paper) available at: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372584>

Douka, V (2018), 'Προσωπικά δεδομένα και εργασιακές σχέσεις: προστασία σύμφωνα με τον Γενικό Κανονισμό Προστασίας Δεδομένων' (Personal data and employment relationships: Protection according to the GDPR) 6 Επιθεώρησης Εργατικού Δικαίου (ΕΕργΔ) (*Labour Law Review*) 641

Douka, V (2020), 'Η λήψη «πλήρως αυτοματοποιημένων αποφάσεων» από τον εργοδότη: Το άρθρο 22 του Γενικού Κανονισμού Προσωπικών Δεδομένων (Κανονισμός 2016/679- Γ.Κ.Π.Δ.)' (Employer's decisions based solely on automated processing: Article 22 of the General Data Protection Regulation (Regulation 2016/679 – GDPR)) 3 Επιθεώρησης Εργατικού Δικαίου (ΕΕργΔ) (*Labour Law Review*) 369

Dreyer, S and Schulz, W (2018), *Was bringt die Datenschutzgrundverordnung für automatisierte Entscheidungssysteme? Potenziale und Grenzen der Absicherung individueller, gruppenbezogener und*

gesellschaftlicher Interessen (What are the benefits of the basic data protection regulation for automated decision systems? Potentials and limits of safeguarding individual, group-related and social interests) (Bertelsmann Stiftung ed) available at: https://algorithmenethik.de/wp-content/uploads/sites/10/2018/04/BSt_DSGVOundADM_dt_ohneCover.pdf

Duarte, F (2018), '5 algoritmos que ya están tomando decisiones sobre tu vida y que quizás tú no sabías' (5 algorithms that are already taking decisions on your life and you may not know it) *BBC* (3 February 2018), available at: www.bbc.com/mundo/noticias-42916502

EELN (2019), *A Comparative Analysis of Gender Equality Law in Europe* (Brussels, European Commission) available at: www.equalitylaw.eu/publications/comparative-analyses

EIGE, 'Gender Impact Assessment' available at: <https://eige.europa.eu/gender-mainstreaming/toolkits/gender-impact-assessment/what-gender-impact-assessment>

ENA (2019), *Rapport collectif sur commande de la mission Etalab, Ethique et responsabilité des algorithmes publics* (Collective report commissioned by the Etalab mission on the Ethics and Responsibility of Public Algorithms) (ENA Year 2018-2019, Molière, June 2019) available at: www.etalab.gouv.fr/wp-content/uploads/2020/01/Rapport-ENA-Ethique-et-responsabilite%C3%A9-des-algorithmes-publics.pdf

Equality Ombudsman (2018), 'Kunskapsöversikt om användningen och utvecklingen av automatiserad databehandling med algoritmer (artificiell intelligens) och stordata och diskriminering eller risker för diskriminering' (Knowledge review of the use and development of automated data processing with algorithms (artificial intelligence) and big data and discrimination or risks of discrimination), Registration No LED 2018/387, Document 55

Equality Ombudsman (2019), 'Promemoria om automatiserad databehandling med algoritmer och risker för diskriminering: inom rekrytering och kreditgivning' (Memorandum on automated data processing with algorithms and risks of discrimination: in recruitment and lending) (Registration No 2018/387, Document 37, 4 October 2019)

eSkills Foundation, 'Analysis of the Gender gap in the Digital Sector in Malta', available at: https://eskills.org.mt/en/analysis%20of%20the%20gender%20gap%20in%20the%20digital%20sector/Documents/Analysis_of_the_gender_gap_in_Maltav2.pdf

Eticas Foundation (2018), 'Algorithmic discrimination. One of the main challenges for social progress in the 21st century' (23 March 2018), available at: <https://eticasfoundation.org/algorithmic-discrimination-one-of-the-main-challenges-for-social-progress-in-the-21st-century/>

European Commission (2019) 'EU countries commit to boost participation of women in digital' (9 April 2019) <https://ec.europa.eu/digital-single-market/en/news/eu-countries-commit-boost-participation-women-digital>

Eurostat (2018), 'Digital economy and society statistics', *Statistics Explained*

Expert Group on Liability and New Technologies – New Technologies Formation (2019), *Liability for Artificial Intelligence and other emerging digital technologies* (European Union)

Fagan, F and Levmore, S (2019) 'The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion' 93 *Southern California Law Review* 11

Farkas, L (2017), *The meaning of racial or ethnic origin in EU law: between stereotypes and identities* (European Commission)

Finanstilsynet (Danish Financial Supervisory Authority) (2019), *Recommendations when using supervised machine learning* (13 September 2019), available at: https://www.dfsa.dk/Supervision/Fintech/Machine_learning_recommendations

Foale, N (2020), 'Back to the Future: How Well Equipped Is Irish Employment Equality Law to Adapt to Artificial Intelligence?' 23(1) *Trinity College Law Review* 170

Fröhlich, W and Spiecker genannt Döhmann, I (2018), 'Können Algorithmen diskriminieren?' (Can algorithms discriminate?) *Verfassungsblog*, (26 December 2018) verfassungsblog.de/koennen-algorithmen-diskriminieren/

Fundacja Edukacyjna Perspektywy (2019) *Women at polytechnics (Kobiety na politechnikach)* available at: www.dziewczynynapolitechniki.pl/pdfy/raport-kobiety-na-politechnikach2019.pdf

Gadhoun Y, Gueyie, J-P and Siala, MK (2007), 'La décision de crédit: Procédure et comparaison de la performance de quatre modèles de prévision d'insolvabilité' (The Credit Decision: procedure and performance comparison of four insolvency forecasting models) 2 (224-225) *La Revue des Sciences de Gestion* 177

Ge, Y and others (2016), 'Racial and gender discrimination in transportation network companies', NBER Working Paper No 22776, available at: <https://www.nber.org/papers/w22776>

Gerards, JH (2019), 'Grondrechten in de platformeconomie' (Fundamental rights in the platform economy) in Gerards, JH and Van Schaik, AC, *Preadvies Vereniging voor Burgerlijk Recht* (Pre-Advice for the Netherlands Association of Private Law) (Uitgeverij Paris, Amsterdam)

Gerards, JH (forthcoming), *Country report: the Netherlands* (Report commissioned by the EU Fundamental Rights Agency as part of its project *Artificial Intelligence, Big Data and Fundamental Rights* (<https://fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights>)) (forthcoming, on file with the author)

Gerards, JH and Zuiderveen Borgesius, F (forthcoming), 'Protected Grounds and the System of Non-Discrimination Law in the Context of Algorithmic Decision-Making and Artificial Intelligence' (in preparation, on file with the authors)

German Women Lawyers' Association (2020), 'Stellungnahme zu dem Weißbuch der EU-Kommission "Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen"' (Statement on the EU Commission White Paper On Artificial Intelligence – A European Concept for Excellence and Trust) (Berlin, 9 June 2020) available at: www.djb.de/fileadmin/user_upload/presse/stellungnahmen/st20-20_WeissbuchKI_EU.pdf

Gillespie, T (2014), 'The Relevance of Algorithms' in Gillespie, T, Boczkowski, PJ and Foot, KA (eds), *Media technologies: Essays on communication, materiality, and society* (Cambridge, MIT Press) 167-194

Gillis, TB and Spiess, JL (2019), 'Big Data and Discrimination' 86 *University of Chicago Law Review* 459

Goddard, K and others (2012), 'Automation bias: a systemic overview of frequency, effect mediators, and mitigators' 19 *Journal of the American Medical Informatics Association* 121

González-Espejo García, MJ (2019), 'Vídeo y conclusiones de las I Jornadas sobre Ética e Inteligencia Artificial: en busca de un algoritmo ético' (Video and conclusions of the 1st Workshop on Ethics and Artificial Intelligence: in search of an ethical algorithm) (Instituto de Innovación Legal, 5 April 2019) available at: <https://institutodeinnovacionlegal.com/jornadas-etica-e-ia/>

- Goodfellow, I, Benjo, Y and Courville, A (2016), *Deep Learning* (Cambridge, MIT University Press)
- Government Office for Science (2016), *Artificial intelligence: opportunities and implications for the future of decision making* (Report GS/16/19, London) available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf
- Grabham, E (2002), 'Law v Canada: New Directions for Equality under the Canadian Charter?' 22 *Oxford Journal of Legal Studies* 641
- De Groen, WP and Maselli, I (2016), *The Impact of the Collaborative Economy on the Labour Market* (Centre for European Policy Studies Special Report)
- Hacker, P (2018), 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic decision-making under EU law' 55 *Common Market Law Review* 1143
- Hajian, S and Domingo i Ferrer, J (2013), 'A Methodology for direct and indirect discrimination prevention in data mining', 25 *IEEE Transactions on knowledge and data engineering* (7) 1445
- Hajian, S and others (2014), 'Generalization-based privacy preservation and discrimination prevention in data publishing and mining', 28 *Data mining and knowledge discovery* (5-6) 1158
- Hajian, S and others (2015), 'Discrimination- and privacy-aware patterns', 29 *Data mining and knowledge discovery* (6) 1733
- Hall, W and Pesenti, J (2017), *Growing the Artificial Intelligence Industry in the UK*, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf
- Hamon R, Junklewitz, H and Sanchez, I (2020), *Robustness and Explainability of Artificial Intelligence* (JRC Technical Report, EU Science Hub) (European Commission, Brussels)
- Hanych, M (2019), 'Umělá inteligencia a lidská práva – nevyhnutelný střet?' (Artificial intelligence and human rights – the inevitable clash) 10 *Bulletin centra pro lidská práva a demokratizaci* 9
- Hatzopoulos, V (2020), 'Η διακυβέρνηση του κυβερνοχώρου' (The governance of cyberspace) in Stephanou, K (ed) (2020), Το σύστημα διακυβέρνησης της Ευρωπαϊκής Ένωσης (The system of governance of the European Union) (Athens, Nomiki Viviothiki) 402
- Haut Conseil à l'Égalité entres les femmes et les hommes (2019), *1er état des lieux du sexisme en France* (1st survey of sexism in France) Rapport No. 2018-01-07 STER 038 available at: www.haut-conseil-egalite.gouv.fr/IMG/pdf/hce_etatdeslieux-sexisme-vf-2.pdf
- Hildebrandt, M (2008), 'Defining profiling: A new type of knowledge?' in Hildebrandt M and Gutwirth S (eds), *Profiling the European citizen* (Springer, Rotterdam) 17
- Hoder, L (2020), 'Jan Lhotský: Umělá inteligencia a lidská práva – obor budoucnosti?' (Jan Lhotský: Artificial intelligence and human rights – a field of the future) *Jiné Právo*, (2 April 2020) available at: <https://jinepravo.blogspot.com/2020/04/jan-lhotsky-umela-inteligence-lidska.html>
- Howard, E (2006), 'The Case for a Considered Hierarchy of Grounds in EU Law' 13 *Maastricht Journal of European and Comparative Law*

Husovec, M (2013), 'Liability for Algorithm Design & Big Data', available at: www.husovec.eu/2013/06/liability-for-algorithm-design-big-data.html

Huq, AZ (2020), 'A Right to a Human Decision' 105 *Virginia Law Review* (forthcoming)

ICO (2017), *Big data, artificial intelligence, machine learning and data protection*, available at: <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

ICO (2020), *Guidance on the AI auditing framework. Draft guidance for consultation*, available at: <https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf>

D'Ignazio, C and Klein, LF (2020), *Data feminism* (MIT Press)

Inspektionen för socialförsäkringen (Swedish Social Insurance Inspectorate) (2018), *Profilering som urvalsmetod för riktade kontroller* (Profiling as a selection method for targeted controls) ISF Report 2018:5, (Stockholm) available at: <https://inspsf.se/publikationer/rapporter/2018/2018-03-26-profilering-som-urvalsmetod-for-riktade-kontroller>

Inspektionen för socialförsäkringen (Swedish Social Insurance Inspectorate) (2018), *Riskbaserade urvalsprofiler och likabehandling* (Risk based selection profiles and equal treatment) Working Report 2018:1AR, (Stockholm) available at: <https://inspsf.se/publikationer/rapporter/2018/2018-06-15-riskbaserade-urvalsprofiler-och-likabehandling>

Institut Montaigne (2020), *Algorithmes : contrôle des biais s.v.p.* (Algorithms: Bias control, please) (March 2020), available at: www.institutmontaigne.org/ressources/pdfs/publications/algorithmes-controle-des-biais-svp.pdf

Izdebski, K (2019), 'Manifest ePaństwa – Postulaty Fundacji ePaństwo na kadencję Sejmu 2019-2023' (The Manifesto – Postulates of the ePaństwo Foundation for the Sejm's term of office 2019-2023) Fundacja ePaństwo, (28 October 2019) available at: <https://epf.org.pl/pl/2019/10/28/manifest-epanstwa-postulaty-fundacji-epanstwo-na-kadencje-sejmu-2019-2023/>

Izdebski, K (2020), '(Nie)słodka tajemnica losowania sędziów' (The (un)sweet secret of the judges' draw) Fundacja ePaństwo, (13 January 2020) available at: <https://epf.org.pl/pl/2020/01/13/nieslodka-tajemnica-losowania-sedziow/>

Jansen, F (2018), 'Data Driven Policing in the Context of Europe', ERC-funded project 'Data Justice: Understanding datafication in relation to social justice' (DATAJUSTICE) starting grant (2018-2023), Working Paper 7 May 2018, 8, available at: www.datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-Data-Driven-Policing-EU.pdf

Jaume-Palásí, L (2017), 'Diskriminierung hängt nicht vom Medium ab' (Discrimination does not depend on the medium), *AlgorithmWatch*, (3 July 2017) available at: <https://algorithmwatch.org/diskriminierung-haengt-nicht-vom-medium-ab/>

Jørgenrud, M (2018), 'Lånekassen tar i bruk kunstig intelligens for å avdekke juks' (The State Educational Loan Fund is using artificial intelligence to detect cheating in students exams) *Digi.no*, (22 January 2018) available at: <http://www.digi.no/artikler/lanekassen-tar-i-bruk-kunstig-intelligens-for-a-avdekke-juks/426299>

Jovović, M (2018), 'Ter Visie – Algoritmische discriminatie' (For Vision – Algorithmic Discrimination) 4 *Arbeid en Onderneming* (Employment and Enterprise) 141

Kamberidou, I, and Pascall, N (2020), 'The digital skills crisis: engendering technology–empowering women in cyberspace', 4 *European Journal of Social Sciences Studies* (6) 1, available at: <http://www.tinyurl.com/ulxl9od>

Kamińska, A (2019), 'Maszyny w służbie państwa. Jak algorytmy wpływają na życie obywateli i obywaterek w Polsce i wybranych krajach regionu' (State-operated machinery. How algorithms affect the lives of citizens in Poland and in the selected countries in the region) Fundacja ePaństwo, (3 May 2019) available at: <https://epf.org.pl/pl/2019/05/13/maszyny-sluzbie-panstwa-algorytmy-wplywaja-zycie-obywateli-obywatelek-polsce-wybranych-krajach-regionu/>

Kantar Emor (2020), 'Klaasseinad ja klaaslagi Eesti IKT-s: müksud naiste osakaalu suurendamiseks IKT valdkonnas hariduses ja tööturul' (Glass walls and glass ceilings in Estonian ICT: opportunities to increase the share of women in ICT in education and the labour market) available at: www.sm.ee/sites/default/files/klaaslagi_ja_klaasseinad_eesti_ikt-s_vahearuanne_2020_kantaremor.pdf

Katyal, SK (2019), 'Private Accountability in the Age of Artificial Intelligence' 66 *UCLA Law Review* 54

Kayser-Bril, N (2019), 'Spain: Legal fight over an algorithm's code', *Algorithm Watch* (12 August 2019), available at: <https://algorithmwatch.org/en/story/spain-legal-fight-over-an-algorithms-code/>

Kerameos, N (2020), 'Γυναίκες στις θετικές επιστήμες: Ένα στοίχημα που πρέπει να κερδίσουμε' (Women in science: A bet that we have to win) available at: <https://www.tovima.gr/2020/03/10/vimagazino/kerameos-gynaikes-stis-thetikes-epistimes-ena-stoixima-pou-prepei-na-kerdisoume/>

Kerikmäe, T (2019), *Autonoomsed intelligentsed tehnoloogiad ja õigusruum* (Autonomous intelligent technologies and legal framework), https://docs.wixstatic.com/ugd/89dea2_c9baeef025a945c2a9585179074fd189.pdf

Khademi, A and others (2019), 'Fairness in Algorithmic Decision Making: An Excursion Through the Lens of Causality' (27 March 2019), available at: <https://arxiv.org/pdf/1903.11719.pdf>

Kilpatrick, C (2014), 'Article 21- Non-Discrimination' in Peers, S and others (eds), *The EU Charter of Fundamental Rights A Commentary* (Hart Publishing)

Kim PT (2017), 'Data-Driven Discrimination at Work' 58 *William and Mary Law Review* 857

Kindsiko, E, Türk, K and Kantšukov, M (2015), *Naiste roll ja selle suurendamise võimalused Eesti IKT sektoris* (Women's role and potential in the ICT sector)

Kirkpatrick, K (2016), 'Battling algorithmic bias: how do we ensure algorithms treat us fairly?' 59 *Communications of the ACM* 16

Kleinberg, J and others (2019), 'Discrimination in the Age of Algorithms', available at: <https://arxiv.org/abs/1902.03731>

Kleinnijenhuis, J (2020), 'Belastingdienst erkent: toch sprake van etnisch profileren' (The Tax and Customs Administration acknowledges: ethnic profiling is happening) *Trouw* (10 May 2020) available at: www.trouw.nl/nieuws/belastingdienst-erkent-toch-sprake-van-etnisch-profileren~b91d1a45/

Knowledge Centre Data & Society (2020), 'Data-Date #2: AI en rekrutering' (Data-Date #2: AI and recruitment) (29 April 2020), available at: <https://data-en-maatschappij.ai/en/news/data-date-2-ai-en-rekrutering>

Koivisto, R and others (2019), *Tekoäly viranomaistoiminnassa – eettiset kysymykset ja yhteiskunnallinen hyväksyttävyys* (Artificial intelligence in authority use – ethical and societal acceptance issues) (Publications of the Government's analysis, assessment and research activities 14/2019, Prime Minister's Office, 1 February 2019) available at: <http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161345/14-2019-Tekoaly%20viranomaistoiminnassa.pdf?sequence=1&isAllowed=y>

Koskenvuo, K (2019), 'Tekoäly avuksi nuorten syrjäytymisen ehkäisemisessä' (AI helpful in preventing youth marginalisation) *Sosiaalivakuutus*, 14 May 2019) available at: <https://sosiaalivakuutus.fi/tekoaly-avuksi-nuorten-syrjäytymisen-ehkäisemisessä/>

Koulu, AR (2018), 'Digitalisaatio ja algoritmit – oikeustiede hukassa?' (Digitalisation and algorithms – Legal science getting lost?) 116(7-8) *Lakimies* 28, 840

Koulu, R and others (2019), *Algoritmi päätöksentekijänä? Tekoälyn hyödyntämisen mahdollisuudet ja haasteet kansallisessa päätöksenteossa* (Algorithm as a decision-maker? The possibilities and challenges of artificial intelligence in the national regulatory environment) (Prime Minister's Office, 2019:44 Helsinki, 25.7.2019) available at: http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161700/TEAS_44_19_Algoritmi%20paatoksenteckijana.pdf

Kroll, JA, Huey, J and others (2017), 'Accountable Algorithms', 165 *University of Pennsylvania Law Review* 633

Kulk, S, Van Deursen, S and others (2020), *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek* (Legal aspects of algorithms that make decisions. An exploratory study) (WODC/ Utrecht University)

Kullmann, M (2018), 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law' 34 *International Journal of Comparative Labour Law and Industrial Relations* 1

Kullmann, M (2019), 'Discriminating job applicants through algorithmic decision-making' *Ars Aequi* 51

Kullmann, M (2019), 'Platformwerk, besluitvorming door algoritmen en bewijs van algoritmische discriminatie' (Platform Work, Algorithmic Decision-Making and Proof of Algorithmic Discrimination) 8 *Ondernemingsrecht* [Business Law]

Külmoja, I (2015), 'Six Myths Keeping Estonian Girls Out of IT', available at: <http://blog.ut.ee/six-myths-keeping-estonian-girls-out-of-it/>

Laas, A (2017), *Practices with Potential* (unpublished national report of Estonia for EIGE study on work-life balance in the ICT sector) available at: <https://eige.europa.eu/publications/work-life-balance/work-life-balance-in-ict>

Ladas, D (2018), Η προσωπικότητα του εργαζομένου (The personality of the employee) (Athens, Nomiki Vivliothiki)

Lambrecht, A and Tucker, C (2019), 'Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads' 65 *Management Science* 2966

Langford, M (2020), 'Taming the Digital Leviathan: Automated Decision-Making and International Human Rights' 114 *American Journal of International Law Unbound* 141 available at: www.cambridge.org/core/journals/american-journal-of-international-law/article/taming-the-digital-leviathan-automated-decisionmaking-and-international-human-rights/5AFE96F03A1B75B63729D60F0F609609

Larus, J and others (2018), 'When Computers Decide: European recommendations on machine-learned automated decision making' (Technical Report Informatics Europe & EUACM, January 2018)

Lazcoz Moratinos, G (forthcoming, on file with the author), 'Modelos algorítmicos, sesgos y discriminación' (Algorithmic models, bias and discrimination)

Leavy, S (2018), 'Gender Bias in Artificial Intelligence: the need for diversity and gender theory in machine learning', 2018 IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE)

Leese, M (2014), 'The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union' 45 *Security Dialogue* 494

Leftheriotou, E (2019), 'Οι Προκλήσεις της Ρομποτικής και της Τεχνητής Νοημοσύνης για το Νομοθέτη και το Δικαστή' (The challenges of Robotics and Artificial Intelligence for the Legislator and the Judge) in Delouka-Igglessi, K, Ligomenou, A and Sinanioti-Maroudi, A (2019), *Δίκαιο και Τεχνολογία* (Law and Technology) (Sakkoulas ed.)

Leong, N and Belzer, A (2017), 'The new public accommodations: race discrimination in the platform economy' 105 *Georgetown Law Journal* 1271

Leslie, D (2019), *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector* (The Alan Turing Institute) available at: www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

Lieder, F and others (2018), 'The anchoring bias reflects rational use of cognitive resources' 25 *Psychonomic Bulletin & Review* 322

Lindma, LM (2019), 'Kaasamise ja mitmekesisuse kiituseks' (In praise of inclusion and diversity), *Director* 5/2019, available at: <https://director.ee/2019/05/02/kaasamise-ja-mitmekesisuse-kiituseks/>

Lobotka, A (2019), *Umělá inteligencia z pohledu antidiskriminačního práva a GDPR* (Artificial intelligence from the point of view of the antidiscrimination law and GDPR) (Wolters Kluwer)

Lundberg, E (2019), *Automated decision-making vs indirect discrimination. Solution or aggravation?* (LLM thesis, Umeå University) available at: www.diva-portal.org/smash/record.jsf?pid=diva2%3A1331907&dsid=-6537

Macron, E (2018), 'Discours du Président de la République sur l'intelligence artificielle' (Presidential speech on artificial intelligence) (29 March 2018) available at: <https://www.elysee.fr/emmanuel-macron/2018/03/29/discours-du-president-de-la-republique-sur-lintelligence-artificielle>

Macron, E (2019) 'Clôture du Global Forum on Artificial Intelligence for Humanity par le Président Emmanuel Macron' (Closing of the Global Forum on Artificial Intelligence for Humanity by President Emmanuel Macron) (30 October 2019) available at: www.elysee.fr/emmanuel-macron/2019/10/30/cloture-du-global-forum-on-artificial-intelligence-for-humanity-par-le-president-emmanuel-macron

Marr, B (2016), 'What Is The Difference Between Deep Learning, Machine Learning and AI?' *Forbes* (8 December 2016) available at: <https://www.forbes.com/sites/bernardmarr/2016/12/08/what-is-the-difference-between-deep-learning-machine-learning-and-ai/#9968fb626cfa>

Martini, M (2017), 'Algorithmen als Herausforderung für die Rechtsordnung' (Algorithms as a challenge for the legal system) 21 *JuristenZeitung* 1017

Matłacz, A (2018), 'Nie można zwalniać za częste L4. Amazon przegrał w sądzie' (One can't lay off for frequent use of L4. Amazon lost in court) (Wyd. WKP, Sekcja: Kadry i BHP, WKP), *Prawo.pl*, (10 December 2018), www.prawo.pl/kadry/zwolnienie-z-pracy-za-czeste-przebywanie-na-l4-wyrok-so-ws,342357.html

Mayson, SG (2018), 'Bias In, Bias Out' University of Georgia School of Law Legal Studies Research Paper No 2018-35

McCrudden, C (2019), *Gender-based positive action in employment in Europe: A comparative analysis of legal and policy approaches in the EU and EEA* (EELN, European Commission) available at: <https://www.equalitylaw.eu/downloads/5008-gender-based-positive-action-in-employment-in-europe-pdf-1-9-mb>

Mercat-Bruns, M (ed) (2020), *Nouveaux modes de détection et de prévention de la discrimination et accès au droit. Action de groupe et discrimination systémique. Algorithmes et préjugés. Réseaux sociaux et harcèlement* (New ways of detecting and preventing discrimination and access to the law. Class action and systemic discrimination. Algorithms and prejudices. Social networks and harassment) (1st edn, Société de législation comparée ed, Collection Trans Europe Experts, Vol 14) available at: www.lgdj.fr/nouveaux-modes-de-detection-et-de-prevention-de-la-discrimination-et-acces-au-droit-9782365170925.html

Mendoza, I and Bygrave, LA (2017), 'The right not to be subject to automated decisions based on profiling' in Synodinou, T and others (eds), *EU Internet Law* (Berlin/Heidelberg, Springer 2017) 77

Mileszyk, N, Paszcza, B and Tarkowski, A (2019), *AlgoPolska. Zautomatyzowane podejmowanie decyzji w służbie społeczeństwu* (AlgoPolska. Automated decision-making at the service of society) (Report 07/2019 by Fundacja Centrum Cyfrowe and Klub Jagielloński, Kraków) available at: <https://centrumcyfrowe.pl/wp-content/uploads/sites/16/2019/12/Algopolska-raport.pdf>

Miller, T (2017), 'Explanation in artificial intelligence: Insights from the social sciences', arXiv Preprint arXiv:1706.07269

Miró Llinares, F (2018), 'Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots' (Artificial Intelligence and criminal justice: beyond the harmful results caused by machines) 20 *Revista de Derecho Penal y Criminología* (Journal of Criminal Law and Criminology) 87

Mitrou, L (2018), 'Data Protection, Artificial Intelligence and Cognitive Services. Is The General Data Protection Regulation (GDPR) "Artificial Intelligence-Proof"? (a study commissioned by Microsoft but expressing the views and opinions of the author), available at: <https://ssrn.com/abstract=3386914>

Mittelstadt, B (2016), 'Auditing for Transparency in Content Personalization Systems' 10 *International Journal of Communication* 4991

Molnar, C (2020), *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, available at: <https://christophm.github.io/interpretable-ml-book/>

Moreira, TC (2019), 'Igualdade de Género no Trabalho 4.0' (Gender equality in 4.0 Work) in Palma Ramalho, MR and Moreira, TC, *A Igualdade nas Relações de Trabalho* (Equality in Employment Relations) (Estudos Apodit 5, AAFDL Editions, Lisbon) 45-68

Moreira, TC (forthcoming), *Algorithms discrimination and Labour Law* (forthcoming in *Anuário de Direitos Humanos* No. 2)

Morelli, C (2018), 'Algoritmi e diritti umani: si rischia la collisione (immaginate a danno di chi?)' (Algorithms and human rights: the risk of collision (imagine to whose detriment?)) *Altalex*, (9 July 2018) available at: <https://www.altalex.com/documents/news/2018/07/09/algoritmi-e-impatto-sui-diritti-umani>

National Delegation to Combat Fraud (Délégation Nationale à la Lutte contre la Fraude) (2014), *Le « data mining », une démarche pour améliorer le ciblage des contrôles* (Data mining: an approach to improve targeting investigations) (Paris, 14 January 2014) available at: www.economie.gouv.fr/files/files/DNLF/fichier_data_mining_joint.pdf

National Equality Ombud (Gleichbehandlungsanwaltschaft), inquiry to the Austrian Labour Market Service, 11 March 2019, see (GER), available at: https://www.gleichbehandlungsanwaltschaft.gv.at/dam/jcr:4dd65631-38ae-43e5-8289-6fbd0cc29ef0/Schreiben_an_AMS_Bundesgesch%C3%A4ftsstelle_final_11.03.2019.pdf

NEN (2020), *Artificiële Intelligentie and Big Data*, available at: www.nen.nl/Normontwikkeling/Doe-mee/Normcommissies-en-nieuwe-trajecten/AI.htm

Nguyen Doan, H T, Briault, C and Moisy, C (2019), 'Artificial intelligence: the future of man?', *Coverstory, Mekkur* (March-April 2019), available at: https://www.cc.lu/uploads/tx_userccpublications/COVER_STORY_OK.pdf

Niklas, J, Sztandar-Sztanderska, K and Szymielewicz, K (2015), *Profiling the unemployed in Poland: social and political implications of algorithmic decision making* (Eliza Jakubiak tr, report by Fundacja Panoptykon, Warsaw) available at: https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf

Nisbett, RE and Ross, L (1980), *Human inference: Strategies and shortcomings of social judgment* (Prentice Hall, Englewood Cliffs)

NLdigital (2019), 'Nieuwe ethische gedragscode AI voor ICT-sector' (New ethics code AI for ICT Sector) (21 March 2019) available at: www.nldigital.nl/news/nieuwe-ethische-gedragscode-ai-voor-ict-sector/

Noble, SU (2018), *Algorithms of oppression: how search engines reinforce racism* (New York University Press)

NOS (2019), 'Algoritmes moeten discriminatie in vacatureteksten opsporen' [Algorithms have to track down discrimination in job advertisements], 10 August 2019. Available at: <https://nos.nl/artikel/2296998-algoritmes-moeten-discriminatie-in-vacatureteksten-opsporen.html>

Numerico, T (2019), 'Algoritmi, strumenti opachi' (Algorithms, opaque instruments) *Il Manifesto*, (30 June 2019) available at: ilmanifesto.it/algoritmi-strumenti-opachi/

Nurmilaakso, T (2017), 'Tekoäly valitsee jo Suomessakin parhaat työntekijät – Tutkijat: "Lisää syrjintää, seksismiä ja rasismia"' (Artificial intelligence in Finland already selects the best employees – Researchers: "Increasing discrimination, sexism and racism") *Yle Finnish Broadcasting Company*, (7 October 2017) available at: <https://yle.fi/aihe/artikkeli/2017/10/07/tekoaly-valitsee-jo-suomessakin-parhaat-tyonhakijat-tutkijat-lisaa-syrjintaa>

NVP (2020), *Sollicitatiecode* (Recruitment Code), available at: www.nvp-hrnetwerk.nl/sollicitatiecode/

Olivi, G (2019), 'Algoritmi che discriminano, ecco perché serve l'approccio giuridico' (Algorithms that discriminate, here is why we need a legal approach) *Network Digital 360*, (4 June 2019) available at: www.agendadigitale.eu/cultura-digitale/algoritmi-che-discriminano-ecco-perche-serve-lapproccio-giuridico/

Van Otterlo, M (2013), 'A machine learning view on profiling', in Hildebrandt M and De Vries K (eds), *Privacy, due process and the computational turn: Philosophers of law meet philosophers of technology* (London, Routledge, 2013) 46

Olsen, H P and others (forthcoming 2021), 'What's in the Box? The legal requirement to explain computationally aided decision-making in public administration' in *Constitutional Challenges in the Algorithmic Society* (Oxford University Press 2021 (forthcoming))

Panagopoulou-Koutnatzi, F (2017), 'Ο Γενικός Κανονισμός για την Προστασία Δεδομένων 679/2016/ΕΕ' (The General Data Protection Regulation 679/2016/EU) (Sakoulas ed.) 112

Panagopoulou-Koutnatzi, F (2017), 'Τα νέα δικαιώματα για τους πολίτες βάσει του Γενικού Κανονισμού Προστασίας Δεδομένων: μια πρώτη αποτίμηση και συνταγματική αξιολόγηση' (The new rights of citizens under the GDPR: a first assessment and constitutional evaluation) *Εφημερίδα Διοικητικού Δικαίου (ΕΔΔ)* (Administrative Law Journal) 81

Papadimitrakis, G (2019), 'Big data και αλγοριθμικές μελέτες επικινδυνότητας – Νέες προκλήσεις στον χώρο της ποινολογίας' (Big data and algorithmic risk research – New challenges in the field of penal sentences) 10 *Ποινική Δικαιοσύνη* (Penal Justice) 1045-1054

Parasuraman, R and Riley, V (1997), 'Humans and automation: use, misuse, disuse, abuse' 39 *Human Factors* 230

Pascual, MG (2019), 'Quién vigila que los algoritmos no sean racistas o sexistas?' (Who monitors that algorithms are not racist or sexist?) *Revista Retina* (17 March 2019) available at: https://retina.elpais.com/retina/2019/03/14/tendencias/1552564034_268678.html

Pasquale, F (2015), *The Black Box Society: The Hidden Algorithms Behind Money and Information* (Harvard University Press)

Pasquale, F (2017), 'Toward a fourth law of robotics: Preserving attribution, responsibility, and explainability in an algorithmic society' 78 *Ohio State Law Journal* 1243

People's Ombudsman (2020), *Izvešće pučke pravobraniteljice za 2019* (Annual Report for 2019) (March 2020, Zagreb) available at: www.ombudsman.hr/wp-content/uploads/2020/03/Izve%C5%A1%C4%87e-pu%C4%8Dke-pravobraniteljice-za-2019.pdf

Pertsinidou, A (2019), 'Η κατάρτιση εξατομικευμένων προφίλ ως εργαλείο αξιολόγησης στο ψηφιακό περιβάλλον εργασίας. Εξόρυξη γνώσης από βάσεις δεδομένων: Κίνδυνοι και προκλήσεις' (Profiling as an evaluation tool in the digital labour environment. Data mining: risks and challenges) 8 *Επιθεώρησης Εργατικού Δικαίου (ΕΕργΔ)* (Labour Law Review) 931

Pessach, D and Shmueli, E (2020), 'Algorithmic Fairness' arXiv:200109784v1 [cs.CY]

Porta, J (2020), 'Algorithmes et risques discriminatoires', in Mercat-Bruns, M (ed) (2020), *Nouveaux modes de detection et de prevention de la discrimination et accès au droit: action de groupe et discrimination systémique, algorithmes et préjugés, réseaux sociaux et harcèlement* (New ways of detecting and preventing discrimination and access to the law. Class action and systemic discrimination. Algorithms and prejudices. Social networks and harassment) (1st edn, Société de législation comparée ed, Collection Trans Europe Experts, Vol 14) available at: www.lgdj.fr/nouveaux-modes-de-detection-et-de-prevention-de-la-discrimination-et-acces-au-droit-9782365170925.html

Pothier, D (2001), 'Connecting Grounds of Discrimination to Real People's Real Experiences' 13 *Canadian Journal of Women and the Law*

Van der Put, M (2019), 'Kan artificiële intelligentie de rechtspraak betoveren?' (Can Artificial Intelligence enchant the judicial branch?) 2 *Rechtstreeks* 50

Quincoces Riesco, A (2018), 'El lado oscuro de los algoritmos' (The dark side of algorithms), *La Vanguardia* (11 June 2018) available at: www.lavanguardia.com/tecnologia/20180611/4518476004/el-lado-oscuro-de-los-algoritmos.html

Rammert, W (2008), 'Where the action is: Distributed agency between humans, machines, and programs', The Technical University Technology Studies Working Papers, TUTS-WP-4-2008

Ranchordas, S (2019), 'Public Values, Private Regulators: Between Regulation and Reputation in the Sharing Economy' 13 *Law & Ethics of Human Rights*

Raso, FA and others (2018), 'Artificial intelligence & human rights: opportunities & risks', The Berkman Klein Center for Internet & Society Research Publication Series: <https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights>

Rathenau Instituut (2020), *Response of the Rathenau Instituut to the European Commission's white paper on Artificial Intelligence* (15 May 2020) available at: www.rathenau.nl/sites/default/files/2020-05/Response%20Rathenau%20Institute%20EC%20Whitepaper%20on%20AI%2015May2020.pdf

Ridgeway, C (2014), 'Why Status Matters for Inequality' 79 *American Sociological Review* 1

Robin-Olivier, S (2012), 'L'émergence de la notion de discrimination indirecte: évolution ou révolution?' [The emergence of the notion of indirect discrimination: evolution or revolution?] in Fines F, Gauthier C and Gautier M (eds), *La non-discrimination entre les européens* (Non-discrimination among Europeans) (Pedone 2012)

Romero, P (2018) 'Manual para evitar "tecnologías machistas"' (Textbook to avoid 'male-chauvinist technologies') *Público* (29 March 2018) available at: www.publico.es/sociedad/cibersociedad-manual-evitar-tecnologias-machistas.html

Romero, P (2019), 'Lo que el algoritmo esconde: matemáticas, sesgos y al final, decisiones humanas' (What the algorithm hides: mathematics, bias and finally, human decisions) *Público* (8 February 2019) available at: www.publico.es/ciencias/inteligencia-artificial-algoritmo-esconde-matematicas-sesgos-final-decisiones-humanas.html

Rovatsos, M, Mittelstadt, B and Koene, A (2019), *Landscape Summary: Bias in Algorithmic Decision-Making. What is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it?* (Centre for Data Ethics and Innovation) available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf

Sacksofsky, U (2010), *Mittelbare Diskriminierung und das Allgemeine Gleichbehandlungsgesetz* (Indirect discrimination and the General Equal Treatment Act) (Berlin)

Saetta, B (2018), 'Algoritmi, intelligenza artificiale, profilazione dei dati: cosa rischiamo davvero come cittadini?' (Algorithms, artificial intelligence, data profiling: what do we really risk as citizens?) *Valigiablu*, (8 January 2018) available at: www.valigiablu.it/algoritmi-dati-rischi/

Ságvári, B (2017), 'Diszkrimináció, átláthatóság és ellenőrizhetőség. Bevezetés az algoritmusetikába' (Discrimination, transparency and control: An introduction of algorithm ethics) 103(3) *Replika* available at: http://replika.hu/system/files/archivum/replika_103-04_sagvari.pdf

Sandberg, A (ed) *Kunskapsöversikter inom det konsumentpolitiska området* (Knowledge reviews in the area of consumer policy), Konsumentverket (Swedish Consumer Agency). 2017:1 available at:

<https://www.konsumentverket.se/globalassets/publikationer/produkter-och-tjanster/gemensamt/rapport-2017-1-antologi-konsumentverket.pdf>

Sandvig, C and others (2014), 'Auditing Algorithms: Research methods for detecting discrimination on internet platforms' Paper presented to *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*

Scheer, J (2019), *Algorithmen und ihr Diskriminierungsrisiko. Eine erste Bestandsaufnahme* (Algorithms and their Risk of Discrimination. A first stocktaking) (State Anti-Discrimination Body of Berlin) available at: www.berlin.de/sen/lads/ueber-uns/materialien/

Schermer, B (2011), 'The limits of privacy in automated profiling and data mining' 27 *Computer Law & Security Review* 45

w(Colloquium Fundamental Rights in the Workplace, 27-28 November 2018, Aarhus University) available at: <https://portal.findresearcher.sdu.dk/en/publications/can-the-use-of-online-targeted-advertising-in-job-adverts-constit>

Schölkopf, B (2019), 'Causality for Machine Learning' (23 December 2019), available at: <https://arxiv.org/pdf/1911.10500.pdf>

Scissons, A (2019), 'How Diversity Can Remove Cultural Bias from Artificial Intelligence', *Thrive Global* (30 August 2019) available at: <https://thriveglobal.com/stories/how-diversity-can-remove-cultural-bias-from-artificial-intelligence/>

Sentis T (2018) *Les enjeux éthiques de la Justice Predictive. Livre* (Ethical Issues in Predictive Justice) (SciencesPo École de Droit and Predictice) available at: www.sciencespo.fr/ecole-de-droit/en/actualites/la-justice-predictive-peut-elle-etre-ethique.html

Skitka, LJ, Mosier, KL and Burdick, M (1999), 'Does automation bias decision-making?' 51 *International Journal of Human-Computer Studies* 991

Skondra, M (2020), 'Αναγνώριση προσώπου και προσωπικά δεδομένα' (Facial recognition and personal data) *Ddikastes* (20 January 2020) available at: www.ddikastes.gr/node/5271

Škop, M and others (2019), *alGOVrithms. State of Play. Report on Algorithms Usage in Government-Citizens Relations in Czechia, Georgia, Hungary, Poland, Serbia and Slovakia* (Krzysztof Izdebski ed, Fundacja ePaństwo) available at: epf.org.pl/pl/wp-content/uploads/sites/2/2019/05/alGOVrithms-State-of-Play-Report.pdf

Sotiropoulos, V (2019), Υπεύθυνος Προστασίας προσωπικών Δεδομένων (Data Protection Officer) (Sakkoulas ed, 2nd edn)

Spanish Agency on Data Protection (2020), *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción* (Adaptation to the GDPR of treatments that incorporate Artificial Intelligence. An introduction) (February 2020) available at: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

Spanish Association for Digitalization (2018), 'Mujeres en la Economía Digital en España' (Women in the digital economy in Spain)

Spielkamp, M (2019), *Automating Society. Taking Stock of Automated Decision-Making in the EU* (A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations,

1st edn, January 2019) available at: https://algorithmwatch.org/wp-content/uploads/2019/01/Automating_Society_Report_2019.pdf

State Anti-Discrimination Body of Berlin (2019), *Algorithmen und das Recht auf digitale Gleichbehandlung. Dokumentation der Fachwerkstatt der Landesstelle für Gleichbehandlung – gegen Diskriminierung* (Algorithms and the right to digital equality. Documentation of the specialist workshop) available at: www.berlin.de/sen/lads/ueber-uns/materialien/

Statistics Finland (2020), *Koulutusalojen sukupuolien mukainen segregaatio* (Gender segregation by field of education), available at: <https://www.stat.fi/tup/tasaarvo/koulutus#segregaatio>

Staudacher, A (2020), 'Datenschutzbehörde dreht den AMS-Algorithmus ab' (Data protection authority turns off the AMS algorithm), *Kurier* (20 August 2020) available at: <https://kurier.at/wirtschaft/datenschutzbehoerde-dreht-den-ams-algorithmus-ab/401006615>

Styczyński, J (2019), 'Algorytmy zamiast działu kadr? RODO studzi zapęły' (Algorithms instead of HR? RODO dampens down the enthusiasm) *Gazeta Prawna* (11 July 2019) available at: <https://serwis.gazetaprawna.pl/praca-i-kariera/artykuly/1421586,rodo-a-automatyczne-systemy-rekrutacji-pracownikow.html>

Surden, H (2019), 'Artificial Intelligence and Law: An Overview', 35 *Georgia State University Law Review* 1306

Swedish Disability Rights Federation (2019), *Förslag för inkluderande AI* (Proposals for inclusive AI), available at: <https://funktionsratt.se/funktionsratt-ratten-att-fungera-i-samhallet-pa-lika-villkor/ett-samhalle-for-alla/1148-2/artificiell-intelligens/forslag-for-inkluderande-ai/>

Symeonidis, I (2019), 'Τεχνητή Νομοσύνη και Δικαιοσύνη' (Artificial Intelligence and Justice), *Εφημερίδα Διοικητικού Δικαίου (ΕΔΔ)* (Administrative Law Journal) 462-469

Szigetvari, A (2020), 'Datenschutzbehörde kippt umstrittenen AMS-Algorithmus' (Data protection authority overturns controversial AMS algorithm), *Der Standard* (20 August 2020) available at: <https://www.derstandard.at/story/2000119486931/datenschutzbehoerde-kippt-umstrittenen-ams-algorithmus>.

Tadelis, S (2016), 'Reputation and Feedback Systems in Oline Platform Markets' 8 *Annual Review of Economics* 321

Tarkowski, A (2019), 'Report. Automating Society. Poland' *AlgorithmWatch*, (29 January 2019) available at: <https://algorithmwatch.org/en/automating-society-poland/>

Tassis, Sp (2018), 'Η εποχή της τεχνητής νομοσύνης' (The era of Artificial Intelligence), *Δίκαιο Μέσω Ενημέρωσης και Επικοινωνίας (ΔΙΜΕΕ)* 4 (Media and Communication Law) 484-494

Teknologirådet (Norwegian Board of Technology) (2015), *Forutseende politi. Kan dataanalyse hjelpe politiet til å være på rett sted til rett tid?* (Foresighted police. Can data analysis help the police to be in the right place at the right time?) available at: https://teknologiradet.no/wp-content/uploads/sites/105/2018/05/ForebyggendeAnalyse_endelig_WEB.pdf

Teknologirådet and NBT (2018), *Kunstig Intelligens – muligheter, utfordringer og en plan for Norge* (AI, possibilities, challenges and a plan for Norway) (Oslo, September 2018) available at: <https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf>

Tennøe, T and Valevatn, J (2018), 'Slik overvåker maskiner følelsene dine' (How machines control your feelings) *Aftenposten*, (26 February 2018, last updated 28 February) available at: www.aftenposten.no/viten/i/ka8jka/slik-overvaaker-maskiner-foelelsene-dine

Þorsteinsson, HF and others (2019), *Iceland and the fourth industrial revolution* (Committee on the Fourth Industrial Revolution, Government of Iceland, Prime Minister's Office, October 2019) available at: www.government.is/lisalib/getfile.aspx?itemid=3fe9f512-eff4-11e9-944e-005056bc530c

Timmer, A (2016), *Gender Stereotyping in the Case Law of the EU Court of Justice* (2016)

Tjong Tjin Tai, E (2017), 'Aansprakelijkheid voor robots en algoritmes' (Liability for robots and algorithms), 14(3) *Nederlands Tijdschrift voor Handelsrecht* (Netherlands Journal for Trade Law) 123

Tobler, C (2005), *Indirect Discrimination: A Case Study into the Development of the Legal Concept under EC Law* (Social Europe Series, Hart Publishing)

Tobler, C (2008), *Limits and potential of the concept of indirect discrimination* (European Commission)

Travlos-Tzanetatos, D (2019), Το εργατικό δίκαιο στην τέταρτη βιομηχανική επανάσταση (Labour Law in the Fourth Industrial Revolution) (Sakkoulas ed)

Tual, M (2018), 'La diversité humaine est un enjeu central pour le développement de l'intelligence artificielle' (Human diversity is a central issue for the development of artificial intelligence), *Le Monde* (30 July 2018) available at: www.lemonde.fr/pixels/article/2018/07/30/intelligence-artificielle-la-diversite-humaine-un-enjeu-central-pour-son-developpement_5337428_4408996.html

Tversky, A and Kahneman, D (1974), 'Judgment under uncertainty: Heuristics and biases' 185 *Science* (4157)

Valonen, T (2020), 'Tiesitkö, että tekoälyyn ja algoritmeihin liittyvää syrjintää valvoo yhdenvertaisuusvaltuutettu?' (Did you know that discrimination related to artificial intelligence and algorithms is monitored by the Non-Discrimination Ombud?) *Yhdenvertaisuusblogi* (non-discrimination blog), (29 April 2020) available at: <https://syrjinta.fi/-/tiesitko-etta-tekoalyyn-ja-algoritmeihin-liittyvaa-syrjintaa-valvoo-yhdenvertaisuusvaltuutettu->

Varshney, KR (2018), 'Introducing AI Fairness 360', *IBM Research Blog*, available at: www.ibm.com/blogs/research/2018/09/ai-fairness-360/

Vervloesem, K (2020), 'In Flanders, an algorithm attempts to make school choice fairer' in Chiusi, F and others (2020) *Automating Society Report 2020* (Algorithm Watch and Bertelsmann Stiftung) 34, available at: <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/10/Automating-Society-Report-2020.pdf>

Verloo, M and Roggeband, C (1996), 'Gender impact assessment: The development of a new instrument in the Netherlands' 14 *Impact Assessment* (1) 3

Vetzo, MJ, Gerards, JH and Nehmelman R (2018), *Algoritmes en grondrechten* (Algorithms and fundamental rights) (Den Haag, Boom Juridisch)

Villani, C (2018), *Donner un sens à l'intelligence artificielle: pour une stratégie nationale et européenne, Report for the Prime Minister* (Giving meaning to artificial intelligence: for a national and European strategy) (8 September 2018), available at: www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

Vorras, A and Mitrou, L (2018), 'Τεχνητή νοημοσύνη και προσωπικά δεδομένα – Μια θεώρηση υπό το πρίσμα του Ευρωπαϊκού Γενικού Κανονισμού Προστασίας Δεδομένων (ΕΕ) 2016/679' (Artificial Intelligence and personal data – An overview under the prism of the European General Data Protection Regulation (EU) 2016/679) 4 Δίκαιο Μέσων Ενημέρωσης & Επικοινωνίας (Δ/ΜΕΕ) 460

Wachter, S (forthcoming), 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' (2020) 35 *Berkeley Technology Law Journal*

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*, available at: <https://ssrn.com/abstract=2903469> or <http://dx.doi.org/10.2139/ssrn.2903469>

Waddington, L and Bell, M (2001), 'More Equal than Others: Distinguishing European Union Equality Directives' 38 *Common Market Law Review*

Waddington, L and Broderick, A (2018), *Combatting disability discrimination and realising equality: A comparison of the UN Convention on the Rights of Persons with Disabilities and EU equality and non-discrimination law* (European Commission and European Network of Legal Experts in Gender Equality and Non-Discrimination)

Wagner, G and Eidenmüller, HGM (2019), 'Down by Algorithms? Siphoning rents, exploiting biases and shaping preferences – the dark side of personalized transactions' 86 *University of Chicago Law Review* 581

Ward, A (2018), 'The Impact of the EU Charter of Fundamental Rights on Anti-Discrimination Law: more a whimper than a bang?' 20 *Cambridge Yearbook of European Legal Studies* 32

Williams, BA, Brooks, CF and Shmargad, Y (2018), 'How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications' 8 *Journal of Information Policy* 78

Wuiame, N (2018), 'Female (un)employment and worklife balance' (Employment Network, Technical dossier No. 8, Transnational ESF Platform)

Xenidis, R (2021), 'Tuning EU equality law to algorithmic discrimination: Three pathways to resilience' 27(6) *Maastricht Journal of European and Comparative Law* 736

Xenidis, R (2020) *Beyond the 'Master's Tools': Putting Intersectionality to Work in European Non-Discrimination Law. A Study of the European Union and the Council of Europe Non-Discrimination Law Regimes* (Florence: European University Institute).

Xenidis, R, 'Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law' (on file with the author)

Xenidis, R (2019), 'Multiple discrimination in EU law: Towards redressing complex inequalities?' in Belavusau, U and Henrard, K (eds), *EU anti-discrimination law beyond gender* (Oxford, Hart Publishing)

Xenidis, R and Senden, L (2019), 'EU Non-discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination' in Bernitz, U and others (eds), *General Principles of EU Law and the EU Digital Order* (Wolters Kluwer 2019)

Yavuz, C (2019), 'Machine Bias: Artificial Intelligence and Discrimination' (LLM thesis, Lund University) available at: <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=8987035&fileId=8987040>

Żebrowski, P (2019), *Kadrowe algorytmy – komputer może dyskryminować pracowników* (Personnel algorithms – computer can discriminate against employees) (Wyd. WKP, Sekcja, Kadry i BHP)

Žliobaitė, I (2017), 'Measuring discrimination in algorithmic decision making' 31 *Data Mining Knowledge Discovery* 1060

Žliobaitė, I and Custers, BHM (2016), 'Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models' 24 *Artificial Intelligence and Law* 183

Zódi, Zs (2017), 'Az információs társadalom legújabb kihívásai a jog számára – Horizontális platformok' (The newest challenges for the law posed by the information society – Horizontal platforms), *Gazdaság és Jog* (No 9) 28

Zuiderveen Borgesius, FJ (2020), 'Strengthening legal protection against discrimination by algorithms and artificial intelligence' *The International Journal of Human Rights* 1

Zuiderveen Borgesius, FJ and Poort, J (2017), 'Online price discrimination and EU data privacy law' *Journal of Consumer Policy* 347

ANNEX – Questionnaire algorithmic discrimination in Europe: challenges and opportunities for gender equality and non-discrimination law

Please answer the questions in this questionnaire by referring to the applicable national legislation and (insofar as relevant) self-regulation, codes of practice or ethical codes, or relevant policy discussions or discussions on legislative proposals, specifying the relevant provisions. If there are national decisions or judgments relating to these questions from any court, quasi-judicial body, human rights institution or national equality body, please include a discussion of the cases and their outcome. Please provide answers to every question separately as well as full references to any source where applicable.

Q1. Examples and problems linked to the impact of algorithms on gender equality and non-discrimination

Existing literature on algorithmic bias flags numerous risks of discrimination in relation to gender. For example, empirical studies reveal that algorithms optimising the distribution of ads on platforms like Facebook might discriminate against women by showing them less job ads for positions in STEM fields and more job ads for positions like supermarket cashiers, thus reinforcing gender segregation in the labour market.⁸⁰⁵ Other examples underline such risks of gender bias in relation to the consumption and supply of goods and services. For instance, Apple is under investigation for sex discrimination in relation to credit offers in the US after the algorithms used to determine credit limits on Apple's credit card provided men with a better credit scores and higher credit caps than women.⁸⁰⁶

a) In your country, does the use of various types of algorithms in the organisation of the labour market (e.g. hiring practices, collaborative economy, platform work...); in the supply of goods and services (e.g. in pricing goods and services, in granting or denying access to goods and services like housing, healthcare, etc., in offering access to certain opportunities like credit...); or in the performance of public administration (e.g. distributing social benefits, combatting fraud, voting, preventing and redressing criminal offences...) pose specific issues in relation to:

- gender equality?
- non-discrimination in relation to other protected grounds?

b) If yes, could you please describe what these problems might be in as much detail as possible? We are interested in knowing whether you, as a national legal expert, think that algorithms will pose new gender equality / non-discrimination problems? Or will the use of AI reinforce existing problems? Or both? And why do you think so? As far as possible, please identify the types of algorithms involved and the relevant stages of algorithmic decision-making (i.e. planning, development and use)?

c) Have you come across examples of cases of algorithmic discrimination in your country (e.g. reported in the national media, in national policy reports or national scientific articles...)? If yes, could you please describe the examples as precisely as possible?

Q2. Specific legal instruments on algorithmic discrimination

Has any specific legal instrument been adopted to combat algorithmic discrimination in your country

805 See Lambrecht, A and Tucker, C (2019), 'Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads' 65 *Management Science* 2966 and Ali, M and others (2019), 'Discrimination Through Optimization: How Facebook's ad Delivery Can Lead to Skewed Outcomes', available at: <https://arxiv.org/abs/190402095>.

806 Kelion, L (2019), 'Apple's 'sexist' credit card investigated by US regulator' (*BBC*, 11 November 2019) available at: www.bbc.co.uk/news/business-50365609.

- in relation to gender equality?
- in relation to other grounds?

a) If yes, what provisions does it make? How is it framed? What is its personal and material scope of application?

b) If not, is there any pending bill or draft legislation in discussion which would aim at addressing algorithmic discrimination specifically?

Q3. Case law on algorithmic discrimination

Have any cases been decided or opinions issued by national courts, tribunals, ombudspersons, equality bodies or other quasi-judicial bodies relevant to algorithmic discrimination?

a) If yes, could you please provide a brief summary of the facts and decision or opinion, as well as any particular legal or other point which you find of relevance?

b) Are there any cases, opinions or decisions that could be relevant to algorithmic discrimination pending in national courts, tribunals, ombudspersons, equality bodies or other quasi-judicial bodies? If yes, and insofar as access is possible, could you please provide a brief summary of the facts and decision, as well as any particular legal or other point which you find of relevance?

c) If no cases have been decided or no relevant opinions issued, do you foresee any difficulties for national judges or other relevant actors in relation to deciding cases of algorithmic discrimination (in particular in relation to establishing responsibility, prima facie evidence, the burden of proof, the qualification of facts as direct or indirect discrimination, sanctions and compensation, etc.)?

Q4. Specific policy instruments and self-regulation on algorithmic discrimination

a) Are there any policy-making efforts ongoing in your country in relation to the fight against algorithmic discrimination (possibly in light of the European Commission's White Paper on AI and the related public consultation)? Have any specific policy measures been adopted or are in discussion? Has the government adopted any specific legal, ethical or policy guidelines? If yes, could you please describe them in detail, explain what their intended effect is and give your own estimation of their efficiency or adequateness?

b) Have any soft regulation instruments to combat algorithmic discrimination (e.g. soft law, political declarations, etc.) been adopted or are they in discussion? If yes, could you please describe them in detail, explain what their intended effect is and give your own estimation of their efficiency or adequateness?

c) Have any self- or co-regulation instruments to combat algorithmic discrimination (e.g. ethical codes, codes of conduct, terms of use, etc.) been adopted by (co-operations of) private companies (including IT platforms, service providers etc.) or are they in discussion? If yes, could you please describe them in detail, explain what their intended effect is and give your own estimation of their efficiency or adequateness?

Q5. Assessment of existing national law: challenges, gaps and weaknesses in relation to algorithmic discrimination

According to you, is the national gender equality and non-discrimination legal framework fit to address the challenges of algorithmic discrimination? Or do you see any gap(s) in the national legal gender equality and non-discrimination framework in your country that would make the (legal) protection against algorithmic discrimination difficult or impossible?

a) If yes, could you please explain what these gaps are (e.g. transposition problem, etc.), what challenges arise from them and provide examples? In particular, we are interested in answers to the following questions:

- In light of the exceptions allowed by EU law regarding the content of media, advertising and education in relation to gender equality and their implementation (or not) in national law, do you think that the law(s) transposing the Gender Goods and Services Directive is/are fit to address algorithmic discrimination?
- In your opinion, does the material and personal scope of national gender equality and non-discrimination law (especially if it goes further than EU law) allow to adequately tackle algorithmic discrimination?
- In your opinion, could national laws transposing EU gender equality law in relation to pregnancy and maternity protection as well as work-life balance in relation to employment play a role in tackling algorithmic discrimination? If so, how?

Algorithms allow to profile people based inter alia on their identity and behaviours, for instance in order to expose them to personalised advertising, realise health-related diagnoses or detect security threats. Algorithmic profiling can rely on numerous categories of data such as gender, age, ethnicity, family status, etc., which it can combine, for example to better target people based on their interests. In 2014, a dating app called OKCupid for instance revealed that black women were more likely to receive substantially worse ratings than other groups of users.⁸⁰⁷ This example is one among many that illustrate how algorithmic profiling poses increased risks of intersectional discrimination.

b) In light of the above, is national gender equality and non-discrimination law well-suited to combat intersectional (and other types of multiple discrimination) arising from algorithms?

c) Have any enforcement problems made the redress of algorithmic discrimination difficult in your country (for example because it was not clear who was to be held responsible and liable for a case of algorithmic discrimination or, for example, because of difficulties to discover, prove or assess discrimination)? If not, do you anticipate any such enforcement problems (including existing enforcement problems, doctrinal problems, interpretation issues in courts, uncertainties about where the liability for algorithmic discrimination lies, etc. that could affect the degree to which the legal framework in place can adequately address algorithmic discrimination)?

d) Beyond gender equality and non-discrimination law, are there any other legal instruments which are or could be directly relevant to the fight against algorithmic discrimination (e.g. legislation or policy on data protection in relation to protected grounds, privacy law, the fight against online discrimination and online gender-based violence, hate speech or harassment, etc.)?

Q6. National legal, political and public debates and discussions

a) Are there any public and/or political discussions/debates on the question of the impact of algorithms on gender equality and non-discrimination law in your national context (e.g. research or policy reports published by public organisations, NGOs, think tanks, private organisations...)? If yes, how do they frame the issue? Could you please describe their main focus and arguments and provide full references? What is your informed opinion about the arguments made by these reports? Do they tackle important questions and do they propose interesting solutions? Do the publication of the AI White Paper of the European Commission and ongoing public consultation play a role in these discussions and debates?

b) The EU General Data Protection Regulation recognises that the processing of personal sensitive data “revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade

⁸⁰⁷ See Hutson, J and others (2018), ‘Debiasing Desire: Addressing Bias & Discrimination on Intimate Platforms’ 2 *Proceedings of the ACM on Human-Computer Interaction*.

union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation" may lead to discrimination and should generally be prohibited.⁸⁰⁸ To what extent is algorithmic discrimination treated through the lens of data protection in national legal, political and public debates and discussions? Do these discussions address the interaction between data protection law and gender equality and non-discrimination law in your national legal context?

c) Are you aware of any legal, political and public debates and discussions or ongoing work on questions of liability and responsibility for algorithmic discrimination? Are there national discussions on who should be held responsible and liable for algorithmic discrimination and who should bear the burden of its prevention (e.g. service providers, algorithm designers, platforms, employers, ...)?

Q7. National legal scholarship

a) Is there specific scholarship on the impact of algorithms for gender equality and non-discrimination law in your national context (e.g. any identifiable scholarly debates on whether national discrimination legislation still 'fits the bill'; what difficulties judges might encounter in front of cases of algorithmic discrimination or the actual difficulties they have faced; how the legislature might address algorithmic discrimination; how the government and other public authorities should go about in anticipating and solving problems of algorithmic discrimination)? If yes, what are the main points/arguments? Do you agree with them and why?

b) In national legal scholarship, to what extent is algorithmic discrimination treated through the lens of data protection in national legal, political and public debates and discussions? Do national legal scholars address the interaction between gender equality and non-discrimination law and data protection law?

c) Is there legal scholarship on questions of liability and responsibility for algorithmic discrimination? Are there scholarly discussions on who should be held responsible and liable for algorithmic discrimination and who should bear the burden of its prevention (e.g. service providers, algorithm designers, platforms, employers, ...)? If yes, what arguments have been put forward and what is your expert opinion on these?

Q8. Relevant institutions and actors

What institutions or other specific actors are, or could be, relevant in the fight against algorithmic discrimination in your national context? What role do/could these institutions play?

Q9. Good practices and monitoring efforts

a) Are you aware of any good practices or recommendations by NGOs, consumer protection organisations, national equality bodies, national human rights institutions or any public or private organisation (e.g. ethical guidelines, monitoring or reporting mechanisms, research and investigation initiatives, victims support infrastructures...)?

b) Are you aware of any monitoring practices/efforts by public or private actors in relation to the impact of algorithms on gender equality and non-discrimination (e.g. either by actors directly involved in the promotion/protection of gender equality and non-discrimination and/or by actors indirectly involved such as data protection agencies, which could for instance issue opinions or reports that could be of relevance to gender equality and non-discrimination law)?

808 Recitals 71 and 75 and Article 9, 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 [2016] OJ L 119/1.

Q10. Female and minority groups' participation in IT and STEM related profession and education

Research shows that the lack of female and minority groups' participation in IT related professions (e.g. machine learning, data science, information technologies, etc.) may lead to the development of algorithmic systems and designs that are biased (e.g. facial recognition software that does not recognise non-white and non-male faces as well as they recognise white and male faces; automated access systems that automatically classify 'Dr.' titles as belonging to male individuals, etc.). Therefore, the participation of women and more broadly non-male groups (transgender, intersex, a-gender, queer people) as well as other minority groups in these professional sectors is key to the development of more equal algorithmic designs and systems.

Are there specific policies in your country, or other relevant initiatives in your national context, that aim to increase female (and non-male) as well as minority groups' participation in education curricula and professions related to IT, data science, etc.?

Q11. Solutions

According to you and in light of your responses above, what are potential solutions to the problems resulting from the use of algorithmic systems delineated above? What good practices that exist in your national context could be used as inspiration elsewhere? Could you provide examples and your own analysis on the adequateness of these solutions?

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at:

https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service: – by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls), – at the following standard number: +32 22999696, or – by email via:

https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from:

<https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

