



European Union Law and Mitigation of Artificial Intelligence-Related Discrimination Risks in the Private Sector: With Special Focus on the Proposed Artificial Intelligence Act

Avrupa Birliği Hukuku ve Özel Sektördeki Yapay Zekâ ile İlişkili Ayrımcılığın Önlenmesi: Taslak Yapay Zekâ Tüzüğü Odağında

Salih Tayfun İnce*

Abstract

Integrating AI systems into decision-making processes in the private sector may place the right to non-discrimination in danger. In order to illustrate this threat, risky fields in the private sector, namely employment, banking, advertising, pricing and insurance, were investigated in this paper with authentic examples of AI-related discrimination. Then, the current EU non-discrimination laws and data protection laws were examined, and it was found out that these EU laws do not have the necessary tools to tackle specific risks arising from AI-related discrimination in the private sector. Therefore, there is an immediate need for new EU legislation equipped with tools which explicitly target AI-related discrimination risks in the private sector. The proposed AI Act may provide new tools against AI-related discrimination in the private sector. Thus, this paper analyzes the proposed AI Act in terms of its potential impacts on mitigating AI-related discrimination risks. Due to the cradle-to-grave approach adopted by the proposed AI Act, providers and users of high-risk AI systems are required to comply with various specific ex-ante and ex-post obligations. It is found out that these obligations can contribute to the mitigation of AI-related discrimination by providing new legal tools. However, these tools are not sufficient in the face of AI-related discrimination risks. Therefore, it is concluded that the crucial need for specific legislation to mitigate AI-related discrimination risks in the private sector is still present.

Keywords

Artificial Intelligence, discrimination, AI-related discrimination, European Union law, non-discrimination laws, data protection law, European Union Artificial Intelligence Act Proposal

Öz

Yapay zeka sistemlerini özel sektördeki karar alma süreçlerine dahil etmek, ayrımcılık yasağının uygulanmasını tehlikeye atabilir. Bu çalışmada, mezkûr tehdidi göstermek için istihdam, bankacılık, reklamcılık, fiyatlandırma ve sigorta gibi özel sektörün bu açıdan riskli alanları, gerçek yaşamdan alınan yapay zeka ile ilişkili ayrımcılık örnekleri ile birlikte incelenmektedir. Ardından, güncel AB ayrımcılık yasağı ve veri koruma düzenlemeleri mercek altına alınmakta ve söz konusu AB düzenlemelerinin özel sektörde yapay zekâ ile ilişkili ayrımcılıktan kaynaklanan risklerle mücadele etmek için gerekli araçlara sahip olmadığı tespit edilmektedir. Bu nedenle, özel sektörde, yapay zekâ ile ilişkili ayrımcılık risklerini yok etmeyi açıkça hedefleyen yeni AB düzenlemelerine acil gereksinim olduğu anlaşılmaktadır. Yapay Zekâ Tüzüğü taslağı, özel sektörde yapay zekâ ile ilişkili ayrımcılığa karşı mücadele için yeni araçlar sağlayabilir. Bu nedenle, bu çalışma, Yapay Zekâ Tüzüğü taslağını, yapay zekâ ile ilişkili ayrımcılık risklerine potansiyel etkisi açısından incelemektedir. Yapay Zekâ Tüzüğü

* **Corresponding Author:** Salih Tayfun İnce (Res. Asst. / Ph.D. Candidate), Fatih Sultan Mehmet Vakıf University, Faculty of Law, Commercial Law Department, İstanbul, Türkiye / Philipps-University of Marburg, Faculty of Law, Marburg, Germany. E-mail: stince@fsm.edu.tr ORCID: 0000-0002-0750-8152

To cite this article: Ince ST, "European Union Law and Mitigation of Artificial Intelligence-Related Discrimination Risks in the Private Sector: With Special Focus on the Proposed Artificial Intelligence Act" Annales de la Faculté de Droit d'Istanbul. Advanced online publication. <https://doi.org/10.26650/annaes.2022.71.0002>



taslađı tarafından benimsenen beşikten mezara yaklaşımı sayesinde, yüksek riskli yapay zekâ sistemlerinin sağlayıcıları ve kullanıcılarının çeşitli öncül ve ardıl yükümlülüklerine uyması gerekmektedir. Bu çalışmada, Yapay Zekâ Tüzüđü taslađı tarafından getirilen bu yükümlülüklerin, yeni yasal araçlar sağlayarak yapay zekâ ile ilişkili ayrımcılığın azaltılmasına katkıda bulunabileceđi tespit edilmiştir. Ancak bu araçlar, yapay zekâ ile ilişkili ayrımcılık riskleri karşısında yeterli değildir. Bu nedenle, özel sektördeki yapay zekâ ile ilişkili ayrımcılık risklerini azaltmak için bu alana özgü hukuki düzenlemelere duyulan yaşamsal gereksinimin varlığını hâlen devam ettirdiđi kanaatine varılmıştır.

Anahtar Kelimeler

Yapay zekâ, ayrımcılık, yapay zekâ ile ilişkili ayrımcılık, Avrupa Birliđi hukuku, ayrımcılık yasađı, veri koruma hukuku, Avrupa Birliđi Yapay Zekâ Tüzüđü Taslađı

I. Introduction

The right to non-discrimination is a fundamental right obtained as a consequence of prolonged and enduring struggles in human history. As a political and legal achievement, it is imperative to protect this right in the face of the emergent risks caused by the utilization of artificial intelligence (“AI”) systems (“**AI-related discrimination**”). However, most of the current laws and regulations against discrimination are not tailored to mitigate AI-related discrimination risks. Hence, the necessity of new interpretational approaches or new legal tools for the prevention of AI-related discrimination risks should be investigated. Since this is a very extensive subject of study, the jurisdictional and sectorial scope of this investigation is restricted in this paper. The European Union (“EU”) law was chosen as the target jurisdiction of this investigation because the EU law is considered as one of the most developed and influential legal systems concerning non-discrimination laws. Moreover, recent efforts of the EU, such as the proposal for the regulation of AI systems, should be investigated since these efforts may provide new tools against AI-related discrimination. The sectorial scope of the investigation is narrowed to the private sector. The underlying reason for this choice is the inadequacy of academic studies regarding AI-related discrimination risks in the private sector, although utilization of AI systems may cause discrimination risks in the private sector as well as the public sector. As a consequence of these reasons, this paper aims to investigate legal safeguards in the EU law against AI-related discrimination risks in the private sector.

To achieve this aim, this paper’s working definition of AI is specified in Part II. Then, the utilization of AI-based decision-making by business organizations is briefly explained. In Part III, discrimination is defined with a legal approach. After that, how AI-related discrimination risks occur is explained with examples. Then, difficulties in tackling these risks are indicated. In Part IV, risky fields for AI-related discrimination in the private sector are investigated. AI-related discrimination risks in employment, banking, advertising, personal pricing and insurance are explained with authentic examples. In Part V, current legal safeguards in EU law against discrimination are reviewed. In particular, EU non-discrimination laws and EU data protection laws are investigated in terms of their suitability for preventing AI-related discrimination risks in the private sector. Afterwards, the European Commission’s Proposal for Artificial Intelligence Act (“**proposed AI Act**”) is examined as to whether it may provide new tools for mitigation of AI-related discrimination risks. In the course of this examination, legal form, preparation and objectives of the proposed AI Act, its risk-based approach, obligations imposed on providers and users of high-risk AI systems, examples regarding the proposed AI Act’s impacts on AI-related discrimination in the private sector and other new legal opportunities provided by the proposed AI Act for mitigation of AI-related discrimination risks in the private sector are addressed.

II. AI And Decision Making

A. AI

AI is considered an umbrella term¹ with many definitions. Adoption of different AI definitions is possible according to the specific goals and contexts of research. Therefore, it is not surprising that there is no consensus on the definition of AI². Nevertheless, a working definition of AI is necessary on which to base subsequent explanations and claims in this paper.

A working definition should be suitable for the intended research and adequately concrete to enable the researcher to work with it directly³. This paper approaches AI from a legal perspective. It attempts to investigate the nature and reciprocal consequences of AI's operation in the non-discrimination law sphere. Since EU law is the target jurisdiction, a legal definition made by EU authorities or bodies should be chosen. The definition of AI provided by the proposed AI Act is adopted in this paper as the most up to date legal definition of AI. The adoption of this definition would also be useful since the impacts of the proposed AI Act in terms of AI-related discrimination in the private sector will be analysed in this paper as well.

The proposed AI Act defines AI as a “*software that is developed with one or more of the techniques and approaches listed in Annex I⁴ and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.*”⁵ This is a conceptual definition of AI with a list of relevant techniques and approaches. In the explanatory memorandum of the proposed AI Act, it is claimed that this definition is clear enough to provide legal certainty and flexible enough to accommodate future developments

- 1 The Royal Society, ‘Explainable AI: the basics’ (Policy briefing) (2019) <royalsociety.org/ai-interpretability> accessed 20 June 2021. Pages 5 and 7.
- 2 Stan Franklin, ‘History, Motivations, and Core Themes’ in Keith Frankish and William W. Ramsey, (eds.), *The Cambridge Handbook Of Artificial Intelligence* (Cambridge University Press, UK 2014) 15; Stephen Lucci and Danny Kopec, *Artificial Intelligence in the 21st Century: A Living Introduction* (Mercury Learning and Information, USA 2016) 4; Max Craglia, (ed.) et al., ‘Artificial Intelligence: A European Perspective’ (Joint Research Centre Working Papers) (2018) 19; The Federal Government of Germany, ‘Artificial Intelligence Strategy’ (2018) <https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf> accessed 20 June 2021. Page 4; Frederik J. Z. Borgesius, ‘Strengthening legal protection against discrimination by algorithms and artificial intelligence’ (2020) 24 10 *The International Journal of Human Rights* 1572, 1573; Sofia Samoli and et al., ‘AI Watch. Defining Artificial Intelligence. Towards An Operational Definition And Taxonomy Of Artificial Intelligence’ (Joint Research Centre Technical Reports) (Luxembourg 2020) 7.
- 3 Pei Wang ‘On the Working Definition of Intelligence’ (1995) <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.5012> accessed 20 May 2021. Page 3.
- 4 “(a) *Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;* (b) *Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;* (c) *Statistical approaches, Bayesian estimation, search and optimization methods.*” Commission (EC) ‘Proposal for a Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts’ COM(2021) 206 final, 21 April 2021. Annex I. (“the proposed AI Act”)
- 5 Article 3 of the proposed AI Act (n 4).

at the same time⁶. In fact, the provision of the list of techniques and methods, which is open to amendments, makes this definition future-proof. However, the extensiveness of its scope may obstruct spot-on identification of AI systems⁷.

The proposed AI Act's AI definition includes machine-learning (“ML”) approaches in its Annex I for the AI techniques and approaches since most of today's groundbreaking developments in the field of AI stem from ML techniques. As a consequence, people mostly use the term ML when they refer to AI⁸. Comprehension of features that separate ML systems from classical computer programming is necessary to understand AI-related discrimination risks explained in the subsequent parts of this paper.

Given ML's role in the development of current AI systems, learning ability is the most distinctive characteristics of ML systems. Classical computers are machines that only execute pre-written instructions (programs) input by humans. After the emergence of ML with learning ability, dynamic decision-making processes that may not be instructed, predicted, or even understood by human designers are developed⁹. These ML systems may produce rules by discovering patterns from the data set. They may apply these rules to new data and may generate original (unexpected) results. As a consequence of this process, ML systems are considered to be based on training instead of explicit programming¹⁰. Since human intervention in and understanding of some ML systems are limited, ML processes may cause problems regarding explainability and transparency. These problems may aggravate detection and mitigation of discrimination risks in AI-based decision-making procedures¹¹.

B. AI-Based Decision-Making

Business organizations are required to make numerous decisions about their investment plans, employment issues, manufacturing, provision of their goods and services, advertisements, etc. Human decision-making can be a long and laboured

6 The proposed AI Act (n 4), 18.

7 It is asserted that this definition covers all information technology systems. Jussi Mäkinen, ‘The EU's AI Regulation Proposal – How to Fix It?’ (2021) <<https://teknologiateollisuus.fi/en/ajankohtaista/eus-ai-regulation-proposal-how-fix-it>> accessed 20 June 2021.

8 Frederik Z. Borgesius, ‘Discrimination, Artificial Intelligence and Algorithmic Decision-making’ (Study Report) (Council of Europe, Strasbourg 2018) <<https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>> accessed 20 June 2021. Page 9.

9 Yavar Bathaee, ‘The Artificial Intelligence Black Box And The Failure Of Intent And Causation’ (2018) 31 2 Harvard Journal of Law & Technology 890, 891; The Royal Society (n 1), 6.

10 The Royal Society, ‘Machine Learning: The Power And Promise Of Computers That Learn By Example’ (Report) (2017) <royalsociety.org/machine-learning> accessed 20 June 2021. Page 16; Craglia et al. (n 2), 20; Blagoj Delipetrev, Chrisa Tsinarakli and Uros Kostić, ‘Historical Evolution of Artificial Intelligence’ (Technical Report) (Publications Office of the European Union, Luxembourg 2020) <https://publications.jrc.ec.europa.eu/repository/handle/JRC120469> accessed 20 June 2021. Page 11.

11 Therefore, assessment and monitoring of the ML systems in terms of discrimination risks should be done according to their dynamic and autonomous features. Wolfgang Wahlster and Christoph Winterhalter, ‘German Standardization Roadmap On Artificial Intelligence’ (German Institute for Standardization and DKE German Commission for Electrical, Electronic & Information Technologies of DIN and VDE) (2020) <<https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungroadmap-en-data.pdf>> accessed 20 June 2021. Page 82.

process if the factors that have to be considered through this process are too much, and relations among these factors are complex. AI systems can shorten and expedite decision-making processes by employing various techniques¹². Additionally, better prediction and resource allocation, improved operations, and personalized digital solutions are some benefits of deploying AI systems. Due to these advantages, important decisions historically taken by humans in various fields, such as health care, public services, justice, farming, education, energy and transportation, security, climate change¹³ and so on, are currently made by AI systems¹⁴.

The utilization of AI in decision-making processes (“**AI-based decision-making**”) may occur in different ways. An AI system may become an assistant for a human to provide information and insights in specific domains when needed. It is called assistant AI¹⁵. Furthermore, humans may constantly collaborate with AI systems to augment the overall intelligence to make decisions on complicated cases¹⁶. These systems are named augmented AI¹⁷. Moreover, AI systems, which have the capacity for autonomous decision-making, may make decisions without human intervention. They are called autonomous AI systems¹⁸. All of these AI systems may carry AI-related discrimination risks at different levels according to the particular features of the systems, such as their autonomy level, intended purpose and place of use. A more detailed explanation regarding how these systems may cause discrimination risks is provided in the next part¹⁹.

III. AI-Related Discrimination

A. Discrimination

It is essential to employ a discrimination definition of the relevant EU laws since this paper is concerned about legal safeguards against AI-related discrimination in the EU laws. Moreover, interpretation of the EU laws is as important as their literal content. Hence, the jurisprudence of the Court of Justice of the European Union

12 Council of Europe, ‘Algorithms and Human Rights: Study On The Human Rights Dimensions Of Automated Data Processing Techniques (In Particular Algorithms) And Possible Regulatory Implications’ (Study Report) (DGI(2017)12) (2018) 26.

13 The proposed AI Act (n 4), 18.

14 The White House, ‘Big Data: Seizing Opportunities, Preserving Values’ (Report) (2014) 64; Joshua A. Kroll et al., ‘Accountable Algorithms’ (2017) 165 University of Pennsylvania Law Review 633, 636.

15 Florian Möslin, ‘Robots in the Boardroom: Artificial Intelligence and Corporate Law’ in Woodrow Barfield, and Ugo Pagallo (eds.), *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar, UK 2018) 649, 657; Kathleen Walch, ‘Is There A Difference Between Assisted Intelligence Vs. Augmented Intelligence?’ *Forbes*. (12 January 2020) <<https://www.forbes.com/sites/cognitiveworld/2020/01/12/is-there-a-difference-between-assisted-intelligence-vs-augmented-intelligence/?sh=3300950426ab>> accessed 20 June 2021.

16 Delipetrev, Tsinarakli and Kostić (n 10), 17.

17 Möslin (n 15), 657; Walch (n 15).

18 Möslin (n 15), 657; Walch (n 15).

19 See Part III-2.

(“CJEU”) plays a decisive role. Although CJEU does not rule on the compatibility of national rules with the EU laws, it officially provides national courts with guidance on interpreting the EU laws in order to decide whether their national laws are compatible with the EU laws. While providing this guidance, the CJEU manifests fundamental principles regarding the definition and properties of discrimination. Additionally, the European Court of Human Rights (“ECtHR”) case law provides guidance for the implementation of the abstract EU laws in concrete cases. Although AI-related discrimination will be examined separately in the next subpart and a more detailed analysis of the EU laws in terms of AI-related discrimination will be provided in part V, definitions of and conceptual approaches to the discrimination in the EU laws and jurisprudence of the CJEU and the ECtHR are briefly provided below.

Discrimination is divided into two main subcategories in the EU laws - direct and indirect discrimination. “(a) *direct discrimination shall be taken to occur where one person is treated less favourably than another is, has been or would be treated in a comparable situation, on any of the grounds referred to in Article 1*”²⁰ according to the EU Employment Equality Directive. The EU Racial Equality Directive further states that

*“Indirect discrimination shall be taken to occur where an apparently neutral provision, criterion or practice would put persons of a racial or ethnic origin 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.”*²¹

As it is evident from the latter definition, the EU law focuses on the discriminatory effects of the treatment/action rather than the intent of the alleged discriminator and the neutrality of the provisions. Likewise, the CJEU considers that

*“indirect discrimination may stem from a measure which, albeit formulated in neutral terms, that is to say, by reference to other criteria not related to the protected characteristic, leads, however, to the result that particularly persons possessing that characteristic are put at a disadvantage”*²².

This neutral criterion or practice that would put persons of a protected class at a particular disadvantage amounts to indirect discrimination and shall be prohibited, “*unless it is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary*”²³. As a concrete example of indirect

20 Article 2(2) of Council Directive (EC) 2000/78 of 27 November 2000 establishing a general framework for equal treatment in employment and occupation [2000] OJ L 303.

21 Article 2(2)(b) of Council Directive (EC) 2000/43 of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L 180. The part of ‘persons of a racial or ethnic origin’ can be replaced with ‘persons from protected classes’ to procure a universal definition for indirect discrimination.

22 See, Case C-83/14, “CHEZ Razpredelenie Bulgaria” AD v Komisia za zashtita ot diskriminatsia, (Judgment of the Court (Grand Chamber) of 16 July 2015, 94.

23 See, Case C-83/14, “CHEZ Razpredelenie Bulgaria” AD v Komisia za zashtita ot diskriminatsia, (Judgment of the Court (Grand Chamber) of 16 July 2015, 111.

discrimination, the CJEU was of the opinion that a department store company which excluded part-time employees from its occupational pension scheme, where that exclusion affected a far greater number of women than men, infringed on the right to equal pay without discrimination based on sex, “*unless the undertaking shows that the exclusion is based on objectively justified factors unrelated to any discrimination on grounds of sex*”²⁴.

In the same vein, the ECtHR considers possible justifications for the investigated treatment and proportionality of the means in its decisions regarding the right to non-discrimination. Hence, a difference in treatment can be qualified as discriminatory “*if it has no objective and reasonable justification, that is, if it does not pursue a legitimate aim or if there is not a reasonable relationship of proportionality between the means employed and the aim sought to be achieved*”²⁵.

In the light of the discrimination definitions provided in EU laws and relevant CJEU and ECtHR precedents, discrimination may occur directly or indirectly. Consideration of the relevant provision’s textual content is prioritized for the direct discrimination assessment. On the other hand, indirect discrimination is assessed by considering its actual effects since the textual content of the provision is already neutral. However, the CJEU and the ECtHR consider certain criteria such as (objective) justification by a legitimate aim, necessity, and proportionality of the means in order to decide whether the investigated practice, which is found indirectly discriminatory, is allowable.

B. AI-Related Discrimination

AI systems may provide societal and economic benefits for society and various industries. However, some AI systems may cause new risks or adverse consequences as well²⁶. AI-related discrimination risk is one of these emergent risks²⁷. Although AI and algorithmic decision-making seem neutral, rational, and unbiased, they may have adverse effects on fundamental rights, including the right to non-discrimination²⁸.

24 See, Case 170/84, *Bilka-Kaufhaus GmbH and Karin Weber von Hartz*, (Judgment of the Court) of 13 May 1986, 31.

25 *Biao v. Denmark (Grand Chamber)* (App no 38590/10) ECHR 24 May 2016, 90.

26 The European Consumer Organization, ‘Automated Decision Making And Artificial Intelligence - A Consumer Perspective’ (Report) (2018) <https://www.beuc.eu/publications/beuc-x-2018-058_automated_decision_making_and_artificial_intelligence.pdf> accessed 20 June 2021. Page 5; Rowena Rodrigues, ‘Legal and human rights issues of AI: Gaps, challenges and vulnerabilities’ (2020) 4 Journal of Responsible Technology <<https://doi.org/10.1016/j.jrt.2020.100005>> accessed 20 June 2021. Page 1; the proposed AI Act (n 4), 1 and 17.

27 Sue Newell and Marco Marabelli, ‘Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of ‘datification’ (2015) 24 Journal of Strategic Information Systems 3, 6; The European Consumer Organization, ‘AI Rights for Consumers’ (Report) (2019) <https://www.beuc.eu/publications/beuc-x-2019-063_ai_rights_for_consumers.pdf> accessed 20 June 2021. Page 5.

28 The White House (n 14), p. 46 and 51; Council of Europe (n 12), p. 26; European Union Agency for Fundamental Rights, ‘#BigData: Discrimination in data-supported decision making’ (Report) (2018) Doi:10.2811/343905, 2; Borgesius (n 8) 7; Philipp Hacker, ‘Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law’ (Pre-print version) (2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3164973&download=yes> accessed 20 June 2021. Page 43; European Union Agency for Fundamental Rights, ‘Getting the Future Right-Artificial Intelligence and Fundamental Rights’ (2020) Doi:10.2811/774118, 68-74; the proposed AI Act (n 4) 21.

In fact, AI systems may carry potential risks for violation of fundamental rights. However, these risks can be actualized with the implementation of these systems by people²⁹. Thus, the terminology of ‘AI-related discrimination’, which makes room for the recognition of human impact on these types of discrimination, is preferred in this paper to refer to discrimination that occurs as a result of the implementation and application of AI systems by people³⁰.

The nature of the AI systems and the fields in which they are utilized requires the application of differentiation techniques such as classifications, clustering, and categorizations. Since AI systems are mostly “black boxes³¹”, AI-based decision-making procedures relying on said differentiation techniques being opaque³². Therefore, it is difficult to detect whether these decisions are based on discriminatory reasons under the guise of differentiation techniques³³. If the discriminatory characteristic of an AI-based decision-making process cannot be proven, legal tools and mechanisms for the protection of the right to non-discrimination cannot be activated. In order to understand underlying reasons for the AI-related discrimination and fulfil this burden of proof, Barocas and Selbst’s explanations regarding which primary steps of AI-based decision-making may lead to discriminatory outcomes can be useful^{34,35}. These steps, which are the definition of target variable and class labels, labelling and collection of training data, using feature selection and proxies, and how they can cause AI-related discrimination risks are explained in the subsequent paragraphs.

29 Jon Kleinberg et al., ‘Discrimination in the Age of Algorithms’ (2018) 10 *Journal of Legal Analysis* 113, 137; Benjamin Wagner et al., ‘Algorithms and human rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications’ (Study Report) (DGI(2017)12) (2018) <<https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>> accessed 20 June 2021. Page 8.

30 There are other denominations in the literature regarding these discrimination risks, such as digital discrimination or algorithmic discrimination. Maddelena Favaretto, Eva De Clercq and Bernice Simone Elger, ‘Big Data and discrimination: perils, promises and solutions. A systematic review’ (2019) 6 *Journal of Big Data* <<https://doi.org/10.1186/s40537-019-0177-4>> accessed 20 June 2021; Raphaële Xenidis, ‘Tuning EU equality law to algorithmic discrimination: Three pathways to resilience’ (2020) 27 6 *Maastricht Journal of European and Comparative Law*. However, this paper prefers the term “AI-related discrimination” since it is wide enough to cover all discrimination risks that have a relation with AI systems.

31 Bathaee (n 9), 893; European Commission’s High-Level Expert Group on Artificial Intelligence, ‘A Definition of AI: Main Capabilities and Disciplines’ (2019) (Report) <<https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>> accessed 20 June 2021. Page 5.

32 Epistemic opacity has been an unavoidable problem in computer science. Users do not (cannot) know every (or epistemically relevant) element of the computer processes between input and output due to their cognitive inadequacy. Paul Humphreys, ‘The philosophical novelty of computer simulation methods’ (2008) 169 *Synthese* 615. <<https://doi.org/10.1007/s11229-008-9435-2>> accessed 20 June 2021. Page 618-619. This problem aggravated with the emergence of new AI techniques such as ML.

33 Borgesius (n 8), 10.

34 Solon Barocas and Andrew D. Selbst, ‘Big Data’s Disparate Impact’ (2016) 104 *California Law Review* 671, 677-694.

35 Similar to Barocas and Selbst (n 34), David Danks and Alex John London, ‘Algorithmic Bias in Autonomous Systems’ (2017) IJCAI’17: Proceedings of the 26th International Joint Conference on Artificial Intelligence 4691-4697 provides a taxonomy of reasons for AI-related discrimination as follows: training data bias, algorithmic focus bias, algorithmic processing bias, transfer context bias and interpretation bias. For the meta-analysis of studies explaining different reasons for AI-related discrimination, see Favaretto, De Clercq, and Elger (n 30), 13. For another taxonomy, see Kleinberg et al. (n 29), 139 ff.

Defining target variables and class labels³⁶ is required for the identification of a problem in AI-based decision-making procedures. For some decisions, such as spam filtering or fraud detection, these definitions are easier due to the binary nature of the categories: spam or not spam, fraud or not fraud. In contrast, some decisions require data engineers to define target variables from scratch and create new classes. For instance, it is exceptionally subjective to determine criteria³⁷ which measures the creditworthiness of a person. In the same vein, the definition of the best candidate for a job is a creative (artistic) task for data engineers as well as employers³⁸. Therefore, subjectivity regarding definitions of target variables and class labels for these types of decisions may cause disproportionate adverse effects on some protected groups³⁹.

Secondly, the features of training data may lead to discrimination in AI decision-making since training data shape the model. Then, the model creates certain outcomes. A data set may be biased because of defective data collection that leads to incorrect, partial, non-representative or over-representative data⁴⁰ or biased data labelling based on biased prior decisions. If the training data is biased as a result of these reasons, the model based on this data will be biased as well.⁴¹ Thus, decisions which are compatible with the right to non-discrimination cannot be expected from these AI systems that are supplied with biased training data⁴².

Another AI decision-making step that can lead to discrimination is feature selection. This is the process of determining categories of data that will be considered (focused, weighted) for the decision-making⁴³. These categories should be determined sufficiently in detail. Otherwise, necessary distinctions cannot be discovered in the process of AI decision-making. This may cause exclusion of certain people because of over-generalization of the representations. For example, AI systems may attribute more value to having graduated from certain reputable colleges instead of more job specific qualifications in recruitment decisions. If disproportionately fewer people from protected classes graduated from these reputable colleges, these people may systemically be excluded in the employment process because of the irrelevant or unelaborate feature selection⁴⁴.

36 The target variable is “*what data miners are looking for.*” And, class labels’ role is to “*divide all possible values of the target variable into mutually exclusive categories.*” Barocas and Selbst (n 34), 678.

37 Barocas and Selbst (n 34), 715.

38 Raphaële Xenidis and Linda Senden, ‘EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination’ in Ulf Bernitz et al. (eds.), *General principles of EU law and the EU digital order* (Kluwer Law International 2020) 151 ff.

39 Barocas and Selbst (n 34), 677-680.

40 Barocas and Selbst (n 34), 684 and 687; European Union Agency for Fundamental Rights (n 28, 2018), 5.

41 Barocas and Selbst (n 34), 681-682; Danks and London (n 35), 4692.

42 A specific manifestation of this reason is the “feedback loop”. These occur when an AI system with learning ability uses its biased outputs as its inputs that generate new biased outputs in future operations. Hence, discrimination is perpetuated and reinforced. Xenidis (n 30), 740.

43 Danks and London (n 35), 4693.

44 Barocas and Selbst (n 34), 688-689.

The existence of proxies for class-membership is yet another problem for AI-based decision-making⁴⁵. Criteria for rational and well-informed decision-making may happen to bring proxies for a class membership since membership of a protected class may be encoded in another piece of data⁴⁶. Consideration of these factors may cause indirect discrimination of people from protected classes, although determination of these criteria aims to make a rational and well-informed assessment⁴⁸. Exclusion of certain factors, which may serve as a proxy for protected classes, from AI-based decision-making process may not guarantee the mitigation of AI-related discrimination⁴⁹. For instance, exclusion of race and ten other variables, which may serve as a proxy for race, does not diminish the discriminatory treatment between different races in terms of credit-riskiness assessment⁵⁰. Therefore, various ex-ante and ex-post requirements⁵¹ should be stipulated by the laws and regulations instead of mere exclusion of certain factors from the AI-based decision-making processes⁵².

Apart from the abovementioned problems, detection and mitigation of AI-related discrimination risks become complicated because of proof requirements, masking techniques, automation bias, technical and legal inabilities, inadequate access to data and code, and the requirement of actual world implementations. These factors are explained in the subsequent paragraphs.

The inclination of AI systems to cause indirect discrimination⁵³ requires closer monitoring and stronger proof since detection and prevention of indirect discrimination is more difficult than direct discrimination. Hence, proof of a seemingly neutral rule or decision, which disproportionately affects a protected group of people, is required. For instance, an AI system may learn from previous loan applications data sets

45 Kleinberg et al. (n 29), 137.

46 It is called "redundant encodings". Cynthia Dwork et al., 'Fairness Through Awareness, Proceedings of the 3rd Innovations in Theoretical Computer Science Conference' (2012) ITCS '12 (Association for Computing Machinery) 214, 226.

47 Devin G. Pope and Justin R. Sydnor, 'Implementing Anti-Discrimination Policies in Statistical Profiling Models' (2011) 3 American Economic Journal: Economic Policy, 206.

48 Barocas and Selbst (n 34), 691-692. For authentic examples of proxy discrimination in different areas of the private sector, see Part IV.

49 Bundesanstalt für Finanzdienstleistungsaufsicht, 'Big data meets artificial intelligence: Challenges and implications for the supervision and regulation of financial services' (2018) (Study) <https://www.bafin.de/SharedDocs/Downloads/EN/dl_bdai_studie_en.html> accessed 20 June 2021. Page 40; European Union Agency for Fundamental Rights (n 28, 2018), 9. It is also claimed that censoring certain data may obstruct the detection of biases and discriminations in AI-based decision-making systems. Betsy Anne Williams et al., 'How Algorithms Discriminate Based on Data they Lack: Challenges, Solutions, and Policy Implications' (2018) 8 Journal of Information Policy 78, 79; European Union Agency for Fundamental Rights (n 28, 2018), 9.

50 Talia B. Gillis and Jan L. Spiess, 'Big Data and Discrimination' (2019) 86 The University of Chicago Law Review 459, 460 and 469.

51 Summary and assessment of methods to mitigate proxy discrimination risks, see Anya E. R. Prince and Daniel Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105 Iowa Law Review 1257, 1300-1317.

52 For a proposed method that reaches a compromise between fairness and predictive accuracy without excluding certain data, see Pope and Sydnor (n 47), 207 ff.

53 Council of Europe (n 12), 27-28; Hacker (n 28), 9 and 11; Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Why Fairness Cannot Be Automated: Bridging The Gap Between EU Non-Discrimination Law And AI' (Draft) (2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3547922> accessed 20 June 2021. Page 45.

that people from certain neighbourhoods are more likely to default on their loan. Therefore, this system may consider postal codes of these areas as negative factors for the prediction of future defaults. If these neighbourhoods correlate with racial origin, it may happen to cause discrimination against people from a protected group⁵⁴.

AI-based decision-making systems may also provide tools which mask intentional discrimination⁵⁵. Users of AI-based decision-making systems may intentionally utilize more complex and remote proxies to get their desired decisions in a discriminatory manner. Therefore, policies against AI-related discrimination should consider intentional and unintentional types of discriminations together⁵⁶.

Automation bias and the technical and legal difficulty of examination of AI systems aggravate the elimination of AI-related discrimination. Wagner argues that *“the human being may often be led to ‘rubber stamp’ an algorithmically prepared decision, not having the time, context or skills to make an adequate decision in the individual case.”*⁵⁷ Moreover, people have a tendency to follow computer-generated advice since they believe that this advice is neutral and unbiased⁵⁸. In fact, this belief causes “automation bias.”⁵⁹ Therefore, if an AI system is utilized in the decision-making process, human intervention or oversight do not always guarantee prevention or rectification of AI-related discrimination.

Detection of AI-related discrimination also requires examination of the code or algorithm while it is processing real world data. This requirement involves two problems. First of all, private organizations tend to withhold their AI systems because of intellectual property, trade secrets or company bylaws⁶⁰. Therefore, it is not easy to examine these AI systems without compulsory regulations. Second, the examination of AI systems in simulated labs does not provide adequate results for complex systems. Thus, these AI systems should be examined while real users are using them⁶¹.

54 Borgesius (n 8), 13.

55 Barocas and Selbst (n 34), 692-693.

56 Nevertheless, AI systems may cause unintentional discrimination more than intentional discrimination because of their complex structure. Borgesius (n 8), 19.

57 Wagner, et al. (n 29), 8.

58 Council of Europe (n 12), 38; European Union Agency for Fundamental Rights (n 28, 2018), 5.

59 Borgesius (n 8), 8 footnote 6. Article 14(4)(b) of the proposed AI Act (n 4).

60 Balázs Bodo et al., ‘Tackling the algorithmic control crisis—the technical, legal, and ethical challenges of research into algorithmic agents’ (2017) 19 Yale Journal of Law & Technology 133, 175; European Union Agency for Fundamental Rights (n 28, 2018), 7; Aaron Rieke, Miranda Bogen and David G Robinson, ‘Public scrutiny of automated decisions: Early lessons and emerging methods’ (Upturn and Omidyar Network Report) (2018) <<https://www.data.govt.nz/assets/Uploads/Public-Scrutiny-of-Automated-Decisions.pdf>> accessed 20 June 2021. Page 19; Borgesius (n 2), 1583; Wachter, Mittelstadt and Russell (n 53), 10.

61 Rieke, Bogen and Robinson (n 60), 19; Susanne Beck et al., ‘Künstliche Intelligenz und Diskriminierung Herausforderungen und Lösungsansätze’ (Whitepaper from Plattform Lernende Systeme) (2019) <<https://www.plattform-lernende-systeme.de/publikationen-details/kuenstliche-intelligenz-und-diskriminierung-herausforderungen-und-loesungsansaetze.html>> accessed 20 June 2021. Page 9.

IV. Selected Risky Fields for AI-Related Discrimination in the Private Sector

A. Employment

The recruitment and placement of employees requires a continuous assessment and selection process. Employers aim to employ the best candidate for the vacant positions. Therefore, they attempt to match the qualifications of the candidates with the requirements of the jobs. After recruitment, employers are required to make decisions regarding the promotion of the staff or termination of employment contracts. For a long time, humans carried out these processes. However, the integration of AI systems into these processes has dramatically increased⁶² because AI systems conduct these processes faster and better than humans. As a consequence, these processes have dramatically transformed with the integration of AI systems⁶³.

Despite the advantages of utilizing AI systems in employment and placement, these systems may cause discrimination because of the underlying systematic reasons explained above⁶⁴. For instance, it was detected that an AI system used by Amazon to screen job applications was discriminating against women. The reason behind this discrimination was the system's self-learning from previous biased data⁶⁵. In other words, the AI system was reinforcing the prior pattern of discrimination.

Sometimes, even the distribution of job advertisements by AI systems may cause discriminatory effects. According to a research experiment, Facebook distributes job advertisements among its users in a way that may cause the ascription of stereotypical affinities to specific groups, such as disproportionate distribution of cashier positions to female users and taxi driver positions to black users. Similarly, another study shows that an AI-based job advertisement system, which is explicitly designed as gender neutral, delivered an advertisement promoting job opportunities in the science, technology, engineering and math fields to men more than women⁶⁶. These distributions may have discriminatory effects, such as exclusion from some parts of the labour market and information, on people from protected classes⁶⁷.

62 Pauline T. Kim, 'Data-Driven Discrimination at Work' (2017) 58 3 William & Mary Law Review 857, 860.

63 According to a study, 72% of the resumes uploaded to commonly used applicant tracking systems (ATS) are eliminated from the recruitment process before a human reviews them. Accesswire, '72% of Resumes are Never Seen by Employers' (16 February 2016) <<https://www.accesswire.com/436847/72-of-Resumes-are-Never-Seen-by-Employers>> accessed 20 June 2021.

64 For underlying reasons of AI-related discrimination risks, see Part III-2.

65 Jeffrey Dastin, 'Amazon scraps secret AI recruiting tool that showed bias against women' *Reuters* (11 October 2018) <<https://www.reuters.com/article/us-amazon-com-jobs-automation-in...-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>> accessed 20 June 2021.

66 Anja Lambrecht and Catherine E. Tucker, 'Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads' (2019) 65 7 Management Science 2966 ff.

67 Muhammad Ali et al., 'Discrimination Through Optimization: How Facebook's ad Delivery Can Lead to Skewed Outcomes' (2019) Proceedings of the ACM on Human-Computer Interaction. <<https://arxiv.org/pdf/1904.02095.pdf>> accessed 20 June 2021. Page 2.

Although criteria for employment may seem rational and reasonable at first glance, applying these criteria by AI systems in employment processes may cause indirect discrimination due to proxies. Therefore, employers who realize these adverse effects may prefer to remove these criteria from their selection process. For instance, it is believed that employee engagement and retention positively correlate with the distance between the home and the workplace. However, some firms do not consider this criterion in their AI-based employment decision-making systems since the distance from the workplace and living in certain neighbourhoods may also serve as a proxy for people from protected classes, especially in the US⁶⁸.

B. Banking

Credit institutions determine whether they will provide credit or how much credit they will provide to a person by assessing the creditworthiness of the relevant person. Income, performance of prior loans, solvency and property ownership are some of the criteria used in creditworthiness assessment. If humans carry out this the assessment, it may take a very long time, as was the case until recently. Additionally, humans' limited cognitive capacity prevents assessment of complex data together. In today's credit market, thousands of complicated transactions occur every second. Hence, credit institutions employ AI systems to assess the creditworthiness of people fast enough to catch up with the speed of the market. Thanks to their velocity, it is evident that AI systems are better at complex data assessment for making credit decisions⁶⁹.

Notwithstanding the advantages of utilization of AI systems in credit transactions mentioned in the previous paragraph, the risk of AI-related discrimination continues to be present⁷⁰. The use of AI-based decision-making systems “*may only shift the locus of discrimination from the bank manager's desk to the programmer's computer screen or to the data scientists' training sets*”⁷¹ since human programmers and data engineers may consciously or unconsciously reflect their biases in the operation and outcomes of AI systems.

It was found in a landmark study that AI-based credit services assign extra interest in mortgage transactions of people from protected classes⁷². Another study conducted

68 Don Peck, 'They're Watching You at Work' *Atlantic* (December 2013) <<http://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681>> accessed 20 June 2021.

69 Bundesanstalt für Finanzdienstleistungsaufsicht, (n 49), 78.

70 Bundesanstalt für Finanzdienstleistungsaufsicht, (n 49), 82. According to a recent survey conducted among global financial sector leaders, “*between 48% and 58% of all respondents believe that mass AI adoption would exacerbate market-level risks*” of privacy breaches, biases and discrimination in financial markets. World Economic Forum, 'Transforming Paradigms A Global AI in Financial Services Survey' (2020) (Report) <http://www3.weforum.org/docs/WEF_AI_in_Financial_Services_Survey.pdf> accessed 20 June 2021. Page 64.

71 Kristin Johnson, Frank Pasquale and Jennifer Chapman, 'Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation' (2019) 88 2 *Fordham Law Review* 499, 506.

72 Robert Bartlett et al., 'Consumer-Lending Discrimination in the FinTech Era' (National Bureau of Economic Research Working Papers 25943) (2019) 21.

in the US asserts that although utilization of AI systems for lending decisions may increase the predictive accuracy of the decision-making procedures, people from certain groups such as Blacks and White Hispanics, are disproportionately less likely to benefit from the adoption of these systems⁷³. Similarly, it is claimed that Apple's credit application algorithm discriminates against women⁷⁴. All of these examples indicate that the utilization of AI-based decision-making in the banking sector has a significant potential for discrimination.

C. Advertisement and Pricing

In capitalist understanding, profit maximization is the primary goal of undertakings. To reach this goal, they have to introduce their goods and services to their target group via advertisement. Additionally, they have to optimize the price of their goods and services considering the supply and demand equilibrium of the specific market. As a result of reaching relevant customers and determining optimum price simultaneously, undertakings maximize their profit. AI-based decision-making systems can be utilized for the improvement of these processes of advertisement and pricing.

The recent dramatic increase in electronic commerce contributes to the success of AI systems by feeding these systems with big data. People's prior shopping transactions, products waiting in their online shopping baskets, their favourite items, their comments and reviews and even how much time they spent on a page of a specific item or how long their mouse cursor stays on a particular product generate big data. Thanks to the processing of these big data by AI systems, undertakings may determine their target group more precisely than ever⁷⁵. For example, Target's AI system can detect pregnant customers (even before these customers find out about their pregnancy!) by reviewing their shopping items and send them pregnancy or baby-related advertisements⁷⁶.

Apart from its benign uses, the utilization of AI systems in advertising and pricing may lead to discriminatory outcomes. For example, research revealed that Facebook

73 Andreas Fuster et al., 'Predictably Unequal? The Effects of Machine Learning on Credit Markets' (2020). SSRN <<https://ssrn.com/abstract=3072038>> accessed 20 June 2021. Page 53-54.

74 Neil Vigdor, 'Apple Card Investigated After Gender Discrimination Complaints' *The New York Times* (10 November 2019) <<https://www.nytimes.com/2019/11/10/business/apple-credit-card-investigation.html>> accessed 20 June 2021; Leo Kelion, 'Apple's 'sexist' credit card investigated by US regulator' *BBC* (11 November 2019) <<https://www.bbc.co.uk/news/business-50365609>> accessed 20 June 2021.

75 The White House (n 14), 40; Newell and Marabelli (n 27), 5.

76 Kashmir Hill, 'How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did' *Forbes* (16 February 2012) <<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=415c8bfe6668>> accessed 20 June 2021; The European Consumer Organization (n 26), 8.

provides advertisers with the opportunity to exclude certain races^{77, 78} and ages⁷⁹ from advertisement target groups. Another research revealed that Google displays better job advertisements to male users than female users⁸⁰. These examples show that AI-based advertisement may debar certain groups of people from certain goods, services, and professional opportunities.

Moreover, the processing of big data regarding customer behaviours may provide predictions regarding customers' reaction to price changes. Hence, AI systems utilized in pricing may classify the customers according to their price sensitivity⁸¹. By using this information, the undertakings can charge different prices for the same product⁸². For example, it was found that a company providing online tutoring services charged different prices to customers from different racial backgrounds. Because of this price differentiation, people with Asian background paid more for the same service⁸³. Another example of AI-related discriminatory pricing is the consideration of a customer's place of residence in pricing. It was claimed that people residing in rural areas are charged more money than people living in large cities for the same product by some online shops in the US⁸⁴. The reason behind this price differentiation is the lack of competition or difficulty to reach physical shops in rural areas. Since rural areas are inhabited by impoverished people, this kind of price differentiation may cause poor people to pay more for the same products on average. Therefore, it is the reinforcement of social inequality and discrimination as well⁸⁵.

77 Julia Angwin, Ariana Tobin and Madeleine Varner, 'Facebook (still) letting housing advertisers exclude users by race' *ProPublica* (21 November 2017) <<https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>> accessed 20 June 2021.

78 The word 'race' is used in this paper to ensure consistency with the terminology of legal sources. It should be noted that the concept of race is a social (human-invented) construct instead of a biologically proven phenomenon.

79 Julia Angwin, Noam Scheiber and Ariana Tobin, 'Dozens of companies are using Facebook to exclude older workers from job ads' *ProPublica* (20 December 2017) <<https://www.propublica.org/article/facebook-ads-age-discrimination-targeting>> accessed 20 June 2021.

80 However, researchers underlined that they could not detect the reasons for these outcomes since the online advertisement system with its various components such as Google, advertisers, websites, and users provide limited visibility for the researchers. Amit Datta, Michael Carl Tschantz and Anupam Datta, 'Automated Experiments on Ad Privacy Settings A Tale of Opacity, Choice, and Discrimination' (2015) Proceedings on Privacy Enhancing Technologies <<https://www.andrew.cmu.edu/user/danupam/dtd-pets15.pdf>> accessed 20 June 2021. Page 1 and 13.

81 The European Consumer Organization, 'The Use Of Big Data And Artificial Intelligence In Insurance' (Report) (2020) <https://www.beuc.eu/publications/beuc-x-2020-039_beuc_position_paper_big_data_and_ai_in_insurances.pdf> accessed 20 June 2021. Page 6.

82 Bundesanstalt für Finanzdienstleistungsaufsicht, (n 49), 42; Borgesius (n 8), 16.

83 Julia Angwin, Surya Mattu and Jeff Larson, 'The tiger mom tax: Asians are nearly twice as likely to get a higher price from Princeton Review'. *ProPublica* (1 September 2015) <<https://www.ProPublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review>> accessed 20 June 2021.

84 Jennifer Valentino-Devries, Jeremy Singer-Vine and Ashkan Soltani, 'Websites vary prices, deals based on users' information' *Wall Street Journal* (24 December 2012) <<https://www.wsj.com/articles/SB1000142412788732377204578189391813881534>> accessed June May 2021.

85 Borgesius (n 8), 36.

D. Insurance

The insurance sector is essentially based on the principle of differentiation between risky and non-risky insureds⁸⁶. In order to predict the risk levels and determine respective premiums, the insurers have always relied on the statistical analysis of data⁸⁷ regarding personal characteristics and habits of insureds, such as smoking, alcohol consumption, age or occupation⁸⁸, and properties of the insured events. These data were in a limited and simple nature in terms of volume and diversity since they were taken directly from the customers. Today, AI systems are fuelled with big data collected through various technological sources⁸⁹. Since the processing of big data provides a larger amount of granular data about the customers, AI systems can perform predictive risk analysis for insurers far better than traditional statistics⁹⁰. Insurers may also determine optimum premiums due to more granular customer segmentation⁹¹. Apart from the core processes of insurance such as risk assessment and price optimization, AI systems can be used in other phases of the insurance process such as advertising, continuous customer communication, claims handling and fraud detection⁹².

Although AI systems may benefit insurers in various ways⁹³, utilization of AI systems in the insurance sector may pose new problems. AI-related discrimination risk is one of these nascent issues⁹⁴. Although discrimination in the insurance sector is a long-standing problem, the AI-based decision-making process may aggravate this problem in terms of its scale and speed⁹⁵. Due to this process, people with certain characteristics and habits may be excluded from insurance coverage because either the insurer does not want to insure these people, or these people cannot afford high premiums⁹⁶. As a result, an AI-based decision-making process with excessive

86 Ronen Avraham, 'Discrimination and Insurance' in Kasper Lippert-Rasmussen (ed.), *The Routledge Handbook of the Ethics of Discrimination*, (Routledge, UK 2018) 335 ff.; Gillis and Spiess (n 50), 459; Borgesius (n 2), 1584.

87 International Association of Insurance Supervisors, 'Issues Paper on the Use of Big Data Analytics in Insurance' (2020) <<https://www.iaisweb.org/page/supervisory-material/issues-papers//file/89244/issues-paper-on-use-of-big-data-analytics-in-insurance>> accessed 20 June 2021. Page 10; Michele Loi and Markus Christen, 'Choosing how to discriminate: navigating ethical trade-offs in fair algorithmic design for the insurance sector' (2021) *Philosophy & Technology* <<https://doi.org/10.1007/s13347-021-00444-9>> accessed 20 June 2021. Page 1.

88 Frederick Schauer, 'Statistical (and non-statistical) discrimination' in Kasper Lippert-Rasmussen (ed.), *The Routledge Handbook of the Ethics of Discrimination*, (Routledge, UK 2018) 42, 51.

89 The White House (n 14), 41; Newell and Marabelli (n 27), 4; European Union Agency for Fundamental Rights (n 28, 2018), 2.

90 Rick Swedloff, 'Risk classification's Big Data (r)evolution' (2014) 21 *Connecticut Insurance Law Journal*, 339, 341; International Association of Insurance Supervisors (n 87), 7.

91 Herb Weisbaum, 'Data Mining Is Now Used to Set Insurance Rates; Critics Cry Foul' *CNBC* (16 April 2014) <<https://www.cnbc.com/2014/04/16/data-mining-is-now-used-to-set-insurance-rates-critics-cry-fowl.html#:~:text=You%20might%20not%20like%20the,to%20provide%20you%20that%20coverage>> accessed 20 June 2021; International Association of Insurance Supervisors (n 87), 8; Loi and Christen (n 87), 2.

92 International Association of Insurance Supervisors (n 87), 14; Loi and Christen (n 87), 2.

93 Swedloff (n 90), 341-342; Bundesanstalt für Finanzdienstleistungsaufsicht, (n 49), 103.

94 The European Consumer Organization (n 81), 3; Loi and Christen (n 87), 23.

95 International Association of Insurance Supervisors (n 87), 12.

96 Bundesanstalt für Finanzdienstleistungsaufsicht, (n 49), 126; International Association of Insurance Supervisors (n 87), 19; The European Consumer Organization (n 81), 3.

fragmentation of customer groups and better predictions may endanger the risk-pooling function of the insurance⁹⁷ by excluding people from insurance coverage⁹⁸.

Direct discrimination can generally be mitigated due to the removal of certain factors such as race, gender or religions of the insurance applicants from the decision-making process. However, indirect discrimination via proxies is a serious problem for AI-based decision-making in the insurance sector since excluding some criteria from the decision-making process does not eliminate indirect discrimination risks. For instance, a journalistic investigation claims, “*men named Mohammed were charged almost £1,000 more for insurance than those called John.*”⁹⁹ Similarly, it was revealed that the postcode of the insurance applicants may dramatically affect the calculation of the insurance premium¹⁰⁰. The UK Financial Conduct Authority discovered that some insurance firms use their or third party datasets “*within their pricing models which may contain factors that could implicitly or potentially explicitly relate to race or ethnicity.*”¹⁰¹ As indicated in these cases, AI-based decision-making systems may calculate different premiums according to the specific information regarding the insurance applicant, such as name and postcode. Since this information may serve as a proxy for the race, gender, or religion of the applicants, people from protected classes may be exposed to AI-related discrimination¹⁰².

Apart from AI-related discrimination caused by proxies, AI-based decision-making systems may reproduce human biases or historical discrimination in the insurance sector. If an AI system is fed with data regarding previous insurance transactions, this AI system may reproduce discriminatory outcomes. For instance, insurers were charging more premiums for African-Americans since race-based premium calculation was not prohibited until the passage of the Civil Rights Act in 1964 in the US¹⁰³. If an AI system uses the last 250 years’ insurance records of the US as the training data, it is an inevitable outcome that this AI system will calculate more premiums for African-American insurance applicants. Even the race of the applicant

97 Borgesius (n 8), 36; The European Consumer Organization (n 81), 8.

98 International Association of Insurance Supervisors (n 87), 8.

99 Ruqaya Izzidien, ‘Higher insurance if you’re called Mohammed? That’s just the start of institutionalized Islamophobia’ *NewStatesman* (23 January 2018) <<https://www.newstatesman.com/politics/uk/2018/01/higher-insurance-if-you-recalled-mohammed-s-just-start-institutionalised>> accessed 20 June 2021; Natalie Corner, ‘Revealed: How your name, your ethnicity, and even your e-mail address could be adding hundreds of pounds to your car insurance premium’ *Dailymail* (29 June 2018) <<https://www.dailymail.co.uk/femail/article-5896561/How-email-address-add-hundreds-car-insurance-premium.html>> accessed 20 June 2021; Harry Kretchmer, ‘Insurers ‘risk breaking racism laws’ *BBC* (9 February 2018) <<https://www.bbc.com/news/business-43011882>> accessed 20 June 2021.

100 Harvey Jones, ‘When the wrong postcode puts insurance out of your reach’ *The Guardian* (30 October 2016) <<https://www.theguardian.com/money/2016/oct/30/wrong-postcode-puts-insurance-out-of-reach>> accessed 20 June 2021.

101 Financial Conduct Authority, ‘Pricing practices in the retail general insurance sector: Household insurance’ (2018) <<https://www.fca.org.uk/publication/thematic-reviews/tr18-4.pdf>> accessed 20 June 2021. Page 15.

102 The European Consumer Organization (n 81), 4 and 12.

103 Jim Probasco, ‘The Insurance Industry Confronts Its Own Racism’ *Investopedia* (1 September 2020) <<https://www.investopedia.com/race-and-insurance-5075141#citation-9>> accessed 20 June 2021.

disregarded in the process, proxies may lead the AI system to designate African-American applicant to charge more premium.

Last but not least, it should be noted that the consequences of AI-generated predictions in the insurance sector are different from the consequences in other sectors, such as medical diagnosis or sentencing¹⁰⁴, since the increase of the predictive accuracy only in the insurance sector may dramatically scale up the AI-related discrimination risks, as it is explained above. Therefore, there should be a trade-off between predictive accuracy and non-discrimination in the insurance sector. Insurers are required to compromise between accuracy and non-discrimination¹⁰⁵ by implementing certain techniques that may mitigate AI-related discrimination risks.

V. Legal Safeguards in European Law Against AI-Related Discrimination

A. Current Laws and Regulations

1. General

General provisions against discrimination exist in different international treaties and conventions that members of the EU are party to¹⁰⁶. Also, Article 14 of the European Convention on Human Rights¹⁰⁷ (“ECHR”) states:

*“The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.”*¹⁰⁸

With similar wording, Article 21(1) of the European Charter of Fundamental Rights (“ECFR”)¹⁰⁹ prohibits

“any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation.”

104 Loi and Christen (n 87), 2.

105 Loi and Christen (n 87), 8.

106 Article 7 of the Universal Declaration of Human Rights, UNGA Res 217/A (10 December 1948); Article 26 of the International Covenant on Civil and Political Rights adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of 16 December 1966 entry into force 23 March 1976, in accordance with Article 49; Article 21 of the Charter of Fundamental Rights of the European Union, 2012/C 326/02 OJ C 326, [2012].

107 European Convention for the Protection of Human Rights and Fundamental Freedoms (1950). <https://www.echr.coe.int/Documents/Convention_ENG.pdf> accessed 20 June 2021.

108 Article 1 of Protocol 12 to the European Convention on Human Rights extends the scope of the protection against discrimination from “rights and freedoms set forth in this Convention” to “any right set forth by law.” Nevertheless, “Article 1 protects against discrimination by public authorities. The article is not intended to impose a general positive obligation on the Parties to take measures to prevent or remedy all instances of discrimination in relations between private persons.” Council of Europe, ‘Explanatory Report to the Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms’ (2000) <<https://rm.coe.int/09000016800cce48>> accessed 20 June 2021. Page 5-6.

109 Charter of Fundamental Rights of the European Union (n 106).

Although the ECFR provides the widest scope for non-discrimination laws (“*any discrimination on any ground*”), it only applies to EU public bodies and member states¹¹⁰. Therefore, relations in the private sector, which this paper focuses on, are not affected directly by these primary laws. Additionally, abstract provisions of these laws regarding the right to non-discrimination require concrete secondary legislation that provides detailed and concrete information about the application.

Secondary sources of EU law, such as EU consumer law, EU competition law and freedom of information law, may be used as legal tools to prevent AI-related discrimination, although these fields are practically untested for this purpose as of yet¹¹¹. On the other hand, EU non-discrimination laws and EU data protection laws may be more promising secondary sources of EU law due to their close connection with the context¹¹². Thus, the subsequent paragraphs discuss whether these laws may provide an efficient legal structure to prevent risks arising from AI-related discrimination.

2. EU Non-discrimination Laws

EU non-discrimination laws mainly consist of four directives, the Racial Equality Directive (2000/43/EC), the Employment Equality Directive (2000/78/EC), the Gender Goods and Services Directive (2004/113/EC) and the Recast Gender Equality Directive (2006/54/EC)¹¹³. There are two main problems regarding the implementation of these directives against AI-related discrimination in the private sector. First of all, the ability to invoke these directives’ is restricted in the private sector. In other words, these directives may be applied to private relations if, and only if, they are transposed into relevant national laws. Secondly, these directives do not contain direct provisions for AI-related discrimination risks because only human discrimination is considered in the preparation phase of these directives¹¹⁴. Due to these problems, general explanations regarding these directives, which are abundant in the literature, are not provided below. Instead of this, the suitability of the current EU non-discrimination laws as tools for mitigating AI-related discrimination is examined, and suggestions for necessary amendments and new legislations are presented in the subsequent paragraphs.

¹¹⁰ Wachter, Mittelstadt and Russell (n 53), 13.

¹¹¹ Borgesius (n 8), 26; Borgesius (n 2), 1582.

¹¹² Borgesius (n 2), 1576.

¹¹³ Although there is a proposal for a horizontal directive that covers all discrimination grounds, it has not been legislated yet. For the proposal see, European Commission, Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation {SEC(2008) 2180} {SEC(2008) 2181} /* COM/2008/0426 final - CNS 2008/0140 */.

¹¹⁴ The European Consumer Organization (n 27), 5; The European Consumer Organization (n 81), 12; Xenidis (n 30), 738. Similarly, the anti-discrimination law in the United States is criticized as well because it does not provide the necessary tools to address AI-related discrimination. Barocas and Selbst (n 34), 694; Kroll et al. (n 14), 636.

Although AI systems generate and process complex and composite classifications, EU non-discrimination law is based on a ‘single axis’ discrimination model. In other words, EU non-discrimination law does not explicitly address intersectional discrimination¹¹⁵. However, the nature of AI-related discrimination is so complex and compounded that it may mostly lead to intersectional discrimination¹¹⁶. Regulations, auditing authorities and de-biasing mechanisms may not catch the AI-related intersectional discrimination if it is not considered explicitly in legislation¹¹⁷. For instance, black women can be victims of intersectional discrimination based on race and gender¹¹⁸. According to a research, around one-fourth of all discrimination in the EU is “*multiple in nature*.”¹¹⁹ Therefore, ideal non-discrimination laws should capture the complex experience of discrimination to protect people from intersectional discrimination caused by the application of AI systems. Indeed, the EU non-discrimination law¹²⁰ is proposed to address intersectional discrimination caused by AI systems¹²¹. Additionally, the Gender Equality Strategy of the EU adapts disaggregation of data as the method to detect intersectional discrimination¹²². For example, a predictive AI system’s accuracy would be evaluated via the partition of the data based on age, gender and intersection of age and gender, as well as the assessment of the aggregated data¹²³¹²⁴.

Sector-specific rules and their *raison d’être* should be considered in the preparation of laws and regulations¹²⁵. Hence, AI-related discrimination risks in different sectors, such as employment, banking, advertising, and insurance, should be separately investigated to reveal the seriousness of the risks. Then, it should be decided whether it is necessary to improve current laws and regulations in the face of AI-related discrimination risks¹²⁶.

115 Xenidis (n 30), 741.

116 Xenidis (n 30), 739.

117 Xenidis (n 30), 741.

118 Xenidis (n 30), 739.

119 European Commission, ‘Special Eurobarometer Discrimination in the EU in 2015’ (Report) (2015).

120 For instance, it is recognized that “*women are often the victims of multiple discrimination*.” Recital 14 of the Council Directive (EC) 2000/43 (n 23).

121 Xenidis (n 30), 738 and 742.

122 “*To get a complete picture of gender-based violence, data should be disaggregated by relevant intersectional aspects and indicators such as age, disability status, migrant status and rural-urban residence*.” European Commission, ‘A Union of Equality: Gender Equality Strategy 2020–2025’ [2020] COM (2020) 152.

123 Xenidis (n 30), 741.

124 However, the CJEU rules that new categories of protected groups cannot be created by the combination of existing categories. For instance, the Court did not assess the combined effect of discrimination based on sexual orientation and age together. Case C443/15, *David L. Parris v Trinity College Dublin and Others* (Judgment of the Court (First Chamber) 24 November 2016).

125 For example, the balance between freedom of contract and protection of consumers should be considered in consumer transactions. *Borgesius* (n 8), 37; *Borgesius* (n 2), 1585.

126 *Borgesius* (n 8), 37.

Moreover, machine learning's reliance on statistical correlations and inferences is not compatible with the essentialist nature of the discrimination concept designated in the EU non-discrimination laws¹²⁷. In order to comply with EU non-discrimination law, membership of a protected class can be disregarded as a factor in AI-based decision-making processes. Although it seems that membership of a protected class is not considered in the decision-making process, other factors may cause discrimination via serving as proxies. For instance, the 'distance to work' criteria in job applications may lead to discrimination since different neighbourhoods –particularly peripheries- can have different racial or socio-economic profiles¹²⁸. Therefore, formalist compliance to EU non-discrimination law does not guarantee mitigation of AI-related discrimination risks, as it is explained above¹²⁹. In order to mitigate proxy-based discrimination risks caused by AI systems, EU non-discrimination laws should stipulate manifold measures against AI-related discrimination risks as well as adapt structural interpretation of the protected grounds instead of an essentialist interpretation¹³⁰.

Lastly, the dynamic nature of the AI systems may lead to new discrimination risks that are not compatible with an exhaustive list of protected classes in the EU secondary non-discrimination laws¹³¹. Race, gender and sexual orientation are protected characteristics that are 'traditionally' protected by non-discrimination laws. However, AI systems with original data processing may invent new classes via differentiation¹³². For example, data processing may differentiate people according to their financial status or price sensitivity¹³³. When these newly invented classes do not match with the characteristics of legally protected groups, non-discrimination laws cannot be applied¹³⁴. Therefore, instead of or in addition to the exhaustive lists, general conditions for establishing protected classes should be determined and incorporated into the non-discrimination laws¹³⁵. Hence, member states may utilize these conditions while they are transposing these EU laws into their national laws.

127 Xenidis (n 30), 745.

128 Peck (n 68); Borgesius (n 8), 29.

129 See Part III-2.

130 Xenidis (n 30), 738 and 748.

131 Borgesius (n 2), 1584; Wachter, Mittelstadt and Russell (n 53), 11; Xenidis (n 30), 738.

132 Scott R. Peppet, 'Regulating the Internet of Things: first steps toward managing discrimination, privacy, security, and consent' (2014) 93 *Texas Law Review* 85, 93; *The White House* (n 14), 53; *Rodrigues* (n 26), 3.

133 Borgesius (n 8), 35.

134 Borgesius (n 8), 35.

135 For a similar suggestion see, Christophe Lacroix, 'Preventing discrimination caused by the use of artificial intelligence' (Report of Committee on Equality and Non-Discrimination at Council of Europe) (2020) 16.

3. EU Data Protection Laws

EU data protection laws may provide some tools to protect people from AI-related discrimination risks. The main sources of data protection laws in the EU are EU's General Data Protection Regulation¹³⁶ (“**GDPR**”) and Council of Europe's Data Protection Convention 108¹³⁷ (“**Modernised Data Protection Convention 108**”).

The most relevant provision of the GDPR that may contribute to the mitigation of AI-related discrimination is Article 22. It prohibits fully automated decision-making, including profiling, if the decision has legal or “*similarly significant effects for the person*.” A court decision or decision about entitlement for social grants or a contract cancellation decision are considered decisions with legal effects for the person¹³⁸. Decisions about employment or credit applications may be considered decisions with similarly significant effects for the person¹³⁹. However, the “*similarly significant effects*” concept is open to discussion. For example, it is not easy to decide whether personal pricing or targeted advertising pass the significant effect threshold. These decisions may pass the said threshold if they bar people from certain goods and services¹⁴⁰.

Decisions that meet the criteria of Article 22, paragraph 1 of the GDPR cannot be taken on a fully automated basis. However, decisions, which are necessary for the performance of or entering into a contract and authorized by EU or member state law to which the controller is subject, and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests or based on the data subject's explicit consent, are regarded as an exception¹⁴¹. Article 22 of the GDPR is criticized by scholars in terms of its effectiveness and comprehensiveness¹⁴². Many automated decisions may be out of the scope of Article 22. Even a superficial intervention (rubber stamping) by a human may bring a decision out of the scope of Article 22¹⁴³. Human intervention in automated decision systems does not guarantee prevention or rectification of AI-related discrimination because of the automation bias¹⁴⁴.

136 Council Regulation (EC) 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.

137 Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data CM/Inf(2018)15-final [2018].

138 Borgesius (n 8), 23. Article 29 of the Data Protection Working Party Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 Wp251rev.01. [2018] 21.

139 Recital 71 of the General Data Protection Regulation 2016/679 (n 136).

140 Article 29 of the Data Protection Working Party Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 Wp251rev.01. [2018] (n 138), 22.

141 Article 22(2) of the General Data Protection Regulation 2016/679 (n 136); Article 29 of the Data Protection Working Party Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 Wp251rev.01. [2018] (n 138), 23.

142 Lilian Edwards and Michael Veale, ‘Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for’ (2017) 16 Duke Law & Technology Review 18, 44; Sandra Wachter, Brent Mittelstadt and Chris Russell, ‘Counterfactual explanations without opening the black box: automated decisions and the GDPR’ (2018) 31 2 Harvard Journal of Law & Technology, 842, 880 ff.

143 Borgesius (n 8), 24.

144 See Part III-2.

Article 9(1)(c) of the Modernised Data Protection Convention 108 enables every individual “to obtain, on request, knowledge of the reasoning underlying data processing where the results of such processing are applied to him or her.” Thus, the Modernised Data Protection Convention 108 provides a more comprehensive “right to explanation” than GDPR since it does not require a decision, which is taken in a fully automated manner and has a significant effect¹⁴⁵.

Another promising tool against AI-related discrimination is the data protection impact assessment (“**DPIA**”) that is required by the GDPR and the Modernised Data Protection Convention 108 for certain situations¹⁴⁶. Article 35, paragraph 1 of the GDPR obliges controllers to conduct a DPIA if “the processing is likely to result in a high risk to the rights and freedoms of natural persons.” Moreover, if an organization takes decisions in a fully automated way and these decisions may “produce legal effects concerning the natural person or similarly significantly affect the natural person”, the performance of a DPIA becomes obligatory according to Article 35 paragraph 3 of the GDPR. Similarly, Article 10(2) of the Modernised Data Protection Convention 108 requires controllers and processors to “examine the likely impact of intended data processing on the rights and fundamental freedoms of data subjects prior to the commencement of such processing.” A DPIA must involve the assessment of unfair or illegal discrimination risks¹⁴⁷. Therefore, the performance of DPIA can be considered an ex-ante tool for the detection and mitigation of AI-related discrimination risks¹⁴⁸.

Despite the existence of some tools against AI-related discrimination, enforcement of data protection laws may not always create an effective result. Data protection laws apply to the processing of personal data when it is related to identifiable persons. Therefore, collective data processing such as predictive modelling is not in the scope of data protection laws¹⁴⁹. Furthermore, the “right to explanation” provided by data protection laws may be of no use because of the lack of meaningful and incomprehensible explanations due to the AI systems’ complexity and black box nature¹⁵⁰. Hence, experts attempt to develop transparency-enhancing systems to provide meaningful and understandable explanations regarding the structure,

145 Borgesius (n 8), 24; Borgesius (n 2), 1580.

146 An impact assessment is defined as “a tool used for the analysis of possible consequences of an initiative on a relevant societal concern or concerns, if this initiative can present dangers to these concerns, with a view to supporting informed decision-making whether to deploy this initiative and under what conditions, ultimately constituting a means to protect these concerns.” Dariusz Kloza et al., ‘Data protection impact assessments in the European Union: complementing the new legal framework towards a more robust protection of individuals’ (d.pia.lab Policy Brief) (2017) <https://cris.vub.be/files/32009890/dpiablab_pb2017_1_final.pdf> accessed 20 June 2021. Page 1.

147 Borgesius (n 8), 22.

148 European Union Agency for Fundamental Rights (n 28, 2018), 8.

149 Borgesius (n 8), 24; Borgesius (n 2), 1581.

150 Jenna Burrell, ‘How the machine ‘thinks’: understanding opacity in machine learning algorithms’ (2016) *Big Data & Society* Doi:10.1177/2053951715622512, 4.

operation, and outcomes of AI systems¹⁵¹. In order to support these efforts, the feature of interpretability may be required for important automated decisions, such as credit applications and employment, taken in the private sector¹⁵², as it is legally required in algorithmic trading¹⁵³.

B. A New Legal Safeguard Against AI-Related Discrimination: The Proposed AI Act

1. Legal Form, Preparation and Objectives of the proposed AI Act in terms of AI-related Discrimination Risks

The current EU laws, including non-discrimination and data protection laws, do not contain express provisions addressing AI-related discrimination risks. Significant problems regarding their transposition, interpretation, and enforcement in terms of AI-related discrimination are explained above. Therefore, they are inadequate to solve problems arising from AI-related discrimination. Although there are self-regulation efforts to prevent AI-related discrimination¹⁵⁴, protection of human rights, e.g. right to non-discrimination, cannot be left to self-regulation. Thus, there is a need for new mandatory legal sources to mitigate AI-related discrimination risks¹⁵⁵.

The choice of legal instrument to regulate emerging technologies has a significant influence on effective enforcement. Therefore, the combination of different legal tools such as statutes and guidelines should be considered. Although statutes contain broad principles consistent with the emerging technologies, their application in practice can be difficult. That being said, publishing guidelines by regulatory authorities can be more useful since regulators can monitor the implementation of these guidelines¹⁵⁶. As a consequence, legislation of a statute and implementation of this statute by the relevant regulatory authority with the help of guidelines may provide better enforcement. Considering these facts, the European Commission introduced its regulation, Proposal for An Act of Artificial Intelligence, on 21 April

151 Mireille Hildebrandt and Serge Gutwirth, 'Concise Conclusions: Citizens Out of Control' in Mireille Hildebrandt and Serge Gutwirth (eds.) *Profiling the European Citizen: Cross-Disciplinary Perspectives* (Springer, 2008) 367.

152 Aaron Rieke, Miranda Bogen, and David G. Robinson, 'Public scrutiny of automated decisions: Early lessons and emerging methods' (Upturn and Omidyar Network Report) (2018) <<https://www.data.govt.nz/assets/Uploads/Public-Scrutiny-of-Automated-Decisions.pdf>> accessed 20 June 2021. Page 6 and 30.

153 Borgesius (n 8), 34.

154 Fairness, Accountability, and Transparency in Machine Learning, <<https://www.fatml.org/>> accessed 20 June 2021; Future of Life Institute, 'The Asilomar AI principles' (2017) <<https://futureoflife.org/ai-principles/>> accessed 20 June 2021; The Montreal Declaration for Responsible AI (2017) <<https://www.montrealdeclaration-responsibleai.com/the-declaration>> accessed 20 June 2021; The Partnership on AI to Benefit People and Society, <<https://www.partnershiponai.org/about/>> accessed 20 June 2021.

155 Christina Angelopoulos et al., 'Study Of Fundamental Rights Limitations For Online Enforcement Through Self-Regulation' (Study Report of Institute for Information Law, University of Amsterdam) (2016) <<https://scholarlypublications.universiteitiden.nl/access/item%3A2869513/view>> accessed 20 June 2021. Page 78; Borgesius (n 8), 27; The European Consumer Organization (n 81), 12.

156 Borgesius (n 8), 33.

2021¹⁵⁷. The legal form of a ‘regulation’ is preferred as the choice of legal instrument since a regulation’s direct applicability can ensure uniform application of new rules in the single (physical and digital) market¹⁵⁸¹⁵⁹. Any guidelines will also be issued according to the proposed AI Act may enhance its enforcement¹⁶⁰.

It is also important to underline that the European Commission affirmed that the ECFR and secondary legislations on non-discrimination, gender equality, consumer production and data protection are considered in the preparation of the proposed AI Act to ensure consistency between these legal sources¹⁶¹.

The proposed AI Act may provide new opportunities and tools for mitigation of AI-related discrimination risks since its precise aim is to ensure the protection and effective enforcement of fundamental rights along with the use and governance of AI systems¹⁶². In particular, the right to non-discrimination and equality between women and men are emphasized as the fundamental rights the protection of which will be promoted and enhanced by the proposed AI Act ¹⁶³.

2. Risk-based Approach

Although certain AI practices are prohibited because of their unacceptable risk¹⁶⁴, the proposed AI Act takes the risk-based approach instead of blanket regulation for all AI systems¹⁶⁵ to proportionally¹⁶⁶ intervene in the market. This legal framework primarily considers the intended purpose and usage area of the AI systems¹⁶⁷ instead of considering these systems exclusively¹⁶⁸. According to this approach, “*AI systems that pose significant risk to the health and safety or fundamental rights of persons*¹⁶⁹”

157 The proposed AI Act (n 4).

158 The proposed AI Act (n 4), 7. The proposed AI Act (n 4) is identified as an inevitable outcome of the Digital Single Market strategy. The proposed AI Act (n 4), 10.

159 As a matter of fact, the scope of territorial application of the proposed AI Act extends beyond the single market since the proposed AI Act comprehends high-risk AI systems deployed even outside of the EU if outcomes of these systems are used in the EU, or they affect natural persons located in the EU. The proposed AI Act (n 4), 20; Article 2(1) of the proposed AI Act (n 4).

160 Commissioner of the Internal Market, Thierry Breton’s comment on the proposed AI Act: “*We will be the first continent where we will give guidelines. So now if you want to use AI applications, go to Europe. You will know what to do and how to do it.*” Javier Espinoza and Madhimuta Murgia, ‘Europe attempts to take leading role in regulating uses of AI’ *Financial Times* (24 April 2021) <<https://www.ft.com/content/360faa3e-4110-4f38-b618-dd695dece90>> accessed 20 June 2021.

161 The proposed AI Act (n 4), 4 and 20.

162 The proposed AI Act (n 4), 3.

163 The proposed AI Act (n 4), 11.

164 Article 1(b) and Article 5 of the proposed AI Act (n 4).

165 The proposed AI Act (n 4), 8.

166 The proposed AI Act (n 4), 7.

167 Article 7(2) of the proposed AI Act (n 4).

168 Jorge Liboreiro, ‘The higher the risk, the stricter the rule’: Brussels’ new draft rules on artificial intelligence’ *Euronews* (21.4.2021) <<https://www.euronews.com/2021/04/21/the-higher-the-risk-the-stricter-the-rule-brussels-new-draft-rules-on-artificial-intelligence>> accessed 20 June 2021.

169 The proposed AI Act (n 4), 3 and 17; Article 7(1)(b) of the proposed AI Act (n 4).

are qualified as high-risk AI systems. Potential risks to fundamental rights, including the right to non-discrimination, protected by the ECFR are considered while classifying the risk level of the AI systems¹⁷⁰. When an AI system is qualified as a high-risk AI system because of its effect on fundamental rights, obligations stipulated in the proposed AI Act will be implemented. Hence, fundamental rights, including the right to non-discrimination, will have an extra level of protection provided by the proposed AI Act.

For non-high-risk AI systems, encouragement and inducement of codes of conduct formation are recommended in the proposed AI Act¹⁷¹. Such codes may be established concerning “*environmental sustainability, accessibility for persons with disability, stakeholder participation in design and development of AI systems and diversity of development teams*.”¹⁷² The proposed AI Act does not prejudice the application of other legislations to non-high-risk AI systems for the mitigation of AI-related discrimination risks. For instance, the General Product Safety Directive¹⁷³ is mentioned in the proposed AI Act as the safety net for non-high-risk AI systems¹⁷⁴.

In conclusion, the proposed AI Act may provide extra tools for mitigation of AI-related discrimination risks since AI systems that may adversely affect fundamental rights, including non-discrimination, are qualified as high-risk AI systems according to the risk-based approach, and detailed obligations are set forth for these systems. In other words, obligations stipulated in the proposed AI Act can be used as tools to prevent AI-related discrimination risks since the scope of the proposed AI Act covers the discriminatory AI systems thanks to the risk-based approach.

3. Obligations of High-risk AI System Providers

The proposed AI Act obliges high-risk AI systems to comply with a group of horizontal compulsory requirements for their trustworthiness¹⁷⁵. Requirements regarding “*quality of data sets used*”¹⁷⁶, “*technical documentation*”¹⁷⁷, and “*record-keeping*”¹⁷⁸, “*transparency and provision of information to users*”¹⁷⁹, “*human oversight*”¹⁸⁰,

170 The proposed AI Act (n 4), 24.

171 Article 69 of the proposed AI Act (n 4); the proposed AI Act (n 4), 9 and 36.

172 Article 69 of the proposed AI Act (n 4); the proposed AI Act (n 4), 16 and 36.

173 Council Directive (EC) 2001/95 of the European Parliament and of the Council of 3 December 2001 on general product safety [2001] OJ L 11.

174 The proposed AI Act (n 4), 37.

175 Chapter 2 of the proposed AI Act (n 4).

176 Article 10 of the proposed AI Act (n 4).

177 Article 11 of the proposed AI Act (n 4).

178 Article 12 of the proposed AI Act (n 4).

179 Article 13 of the proposed AI Act (n 4).

180 Article 14 of the proposed AI Act (n 4).

and robustness, accuracy, and cybersecurity^{181,7} are indicated as mandatory requirements for high-risk systems to mitigate risks against fundamental rights¹⁸².

As evident from explanations in the following paragraphs, most of the obligations stipulated in the proposed AI Act are addressed to providers of AI systems¹⁸³. Therefore, without considering the designer or the developer of the system, the provider¹⁸⁴ (or the manufacturer if the system is not placed on the market or put into service independently from the manufacturer's product) is obliged to bear the liability arising from non-compliance with obligations set forth in the proposed AI Act regarding placement of the high-risk AI system into the market or putting it into service¹⁸⁵. Determination of liability structure for the use of high-risk AI systems may incentivize liable persons or organizations to take necessary measures, such as measures against discrimination risks.

The proposed AI Act contains ex-ante obligations such as testing and risk management. Moreover, ex-post obligations, such as post-market monitoring and bias monitoring, provide transparency and traceability for the effectiveness of redress mechanisms for affected persons¹⁸⁶.

As a part of ex-ante obligations, high-risk AI systems are required to have CE marking according to the proposed Act¹⁸⁷. Acquiring the procedure of CE marking¹⁸⁸ will guarantee the conformity of AI systems with all relevant EU-wide harmonized standards¹⁸⁹. Moreover, conformity assessment for high-risk AI systems should be performed before their placement on the market¹⁹⁰ or entering into service^{191,92}. Compliance of AI systems with obligations of the proposed AI Act should be evaluated in the conformity assessment¹⁹³. For some of the AI systems listed in Annex III (2-8), providers may conduct conformity assessments without the involvement of

181 Article 15 of the proposed AI Act (n 4).

182 The proposed AI Act (n 4), 29.

183 Specific criteria are set by Article 28 of the proposed AI Act (n 4) for the consideration of distributors, importers, users or any other third parties as providers.

184 The provider is defined in Article 3(2) of the proposed AI Act (n 4) as "*a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge.*"

185 The proposed AI Act (n 4), 31.

186 The proposed AI Act (n 4), 11.

187 The proposed AI Act (n 4), 33.

188 For the procedure of CE marking, see European Commission, 'Internal Market, Industry, Entrepreneurship and SMEs' <https://ec.europa.eu/growth/single-market/ce-marking/manufacturers_en> accessed 20 June 2021.

189 For the list of harmonised standards, see European Commission, 'Harmonised Standards' <https://ec.europa.eu/growth/single-market/european-standards/harmonised-standards_en> accessed 20 June 2021.

190 Placing on the market is defined in Article 3(9) of the proposed AI Act (n 4) as "*the first making available of an AI system on the Union market.*"

191 Putting into service is defined in Article 3(11) of the proposed AI Act (n 4) as "*the supply of an AI system for first use directly to the user or for own use on the Union market for its intended purpose.*"

192 The proposed AI Act (n 4), 32.

193 The proposed AI Act (n 4), 32.

third parties such as independent experts or notified bodies. In other words, much of high-risk AI systems' compliance with provisions of the proposed AI will be ensured through self-assessment. Although the self-assessment approach will be welcomed by the AI providers¹⁹⁴, it would be better either to restrict these categories requiring self-assessment or to oblige providers to get conformity assessments by third parties¹⁹⁵. Otherwise, conformity assessment will simply be a tick-box procedure for many of the AI systems¹⁹⁶. Hence, it may neither provide a reliable enforcement environment for the conformity assessment stipulated in the proposed AI Act nor create adequate public trust toward AI systems that is also aimed at the proposed AI Act.

Substantial modification of the AI systems requires a new conformity assessment. The proposed AI Act recognizes the 'learning' feature of the AI systems by exempting changes to the algorithm and its performance due to 'learning' if rules for change via learning are pre-determined by the provider and evaluated in the conformity assessment¹⁹⁷. Apart from this exemption, if it is realized that the system is no longer in conformity with the proposed AI Act, the provider is required to take appropriate corrective actions such as bringing that system into conformity or withdrawing or recalling the system¹⁹⁸. The ex-ante obligations of risk and conformity assessment and CE marking may promote the compliance-by-design approach¹⁹⁹. Hence, it may help early detection and mitigation of AI-related discrimination risks.

The proposed AI Act requires continuous human oversight for the use of high-risk AI systems²⁰⁰. Prevention or minimization of the risks to health, safety or fundamental rights is indicated as the aim of human oversight. Individuals assigned to human oversight tasks should have the necessary information and experience to fulfil their obligations. In fact, the proposed AI Act explicitly states that individuals assigned to human oversight roles have to pay attention to automation bias to prevent the risk of overreliance on the outcomes of AI systems. Although human oversight does not guarantee the prevention of AI-related discrimination, it may help early detection and mitigation of AI-related discrimination risk by providing an extra filter.

194 Natalie Pettinger-Kearney et al., 'The EU's proposed AI Regulation' (2021) <<https://www.freshfields.com/en-gb/our-thinking/campaigns/digital/artificial-intelligence/the-eus-proposed-ai-regulation/>> accessed 20 June 2021.

195 Sarah Chander and Ella Jakubowska, 'EU's AI law needs major changes to prevent discrimination and mass surveillance' *EDRI* (2021) <<https://edri.org/our-work/eus-ai-law-needs-major-changes-to-prevent-discrimination-and-mass-surveillance/>> accessed 20 June 2021.

196 Mark MacCarthy and Kenneth Propp, 'Machines Learn That Brussels Writes the Rules: The EU's New AI Regulation' *Lawfare Blog* (28 April 2021) <<https://www.lawfareblog.com/machines-learn-brussels-writes-rules-eus-new-ai-regulation>> accessed 20 June 2021.

197 The proposed AI Act (n 4), 33.

198 Article 21 of the proposed AI Act (n 4).

199 Friederike Reinhold and Angela Müller, 'Algorithm Watch's response to the European Commission's proposed regulation on Artificial Intelligence – a major step with major gaps' (Report) (22 April 2021) <<https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021/>> accessed 20 June 2021.

200 Article 14 of the proposed AI Act (n 4).

As is explained above, the opacity and black-box nature of the AI systems restrain people from learning and challenging the discriminatory outcomes of AI-based decision-making procedures. The proposed AI Act sets forth the establishment of stand-alone high-risk AI systems registry²⁰¹, where the providers will submit meaningful information²⁰² regarding features and conformity assessment results of AI systems²⁰³. Thus, transparency toward the public would be increased²⁰⁴. Thanks to this registration, people affected by these systems may get more information about relevant AI systems. Information provided in the registration may help these affected people challenge these AI systems' outcomes in terms of non-discrimination laws.

As an ex-post obligation, post-market monitoring is stipulated in the proposed AI Act. Preparation of post-market monitoring plan and performance of post-market monitoring²⁰⁵ by the providers is required for high-risk AI systems²⁰⁶ to mitigate risks arising from AI systems. Thus, the providers are required to inform the relevant authorities if the use of their systems leads to a violation of national or EU law protecting fundamental rights, including the right to non-discrimination²⁰⁷. Another ex-post obligation is bias monitoring, which involves the examination of training, validation and testing data and operation of AI systems regarding possible biases, which is set forth in the proposed AI Act²⁰⁸²⁰⁹. Both of these ex-post obligations provide very useful tools for the detection and mitigation of AI-related discrimination risks because these obligations are compatible with the learning feature of the emergent AI systems. Thanks to these ex-post obligations, unpredictable changes in the outcomes of an AI-based decision-making system would be monitored in terms of whether these changes cause AI-related discrimination risks.

For the entire life cycle of high-risk AI systems, clear obligations to ensure compliance with fundamental rights are stipulated²¹⁰. The establishment and maintenance of risk management systems throughout the entire life cycle of high-

201 Article 13(3) of the proposed AI Act (n 4).

202 For the information to be submitted upon the registration of high-risk AI systems, see Annex VIII of the proposed AI Act (n 4)

203 The proposed AI Act (n 4), 12.

204 The proposed AI Act (n 4), 33.

205 The post-market monitoring is defined in Article 3(25) of the proposed AI Act (n 4) as *“all activities carried out by providers of AI systems to proactively collect and review experience gained from the use of AI systems they place on the market or put into service for the purpose of identifying any need to immediately apply any necessary corrective or preventive actions.”*

206 Article 61 of the proposed AI Act (n 4).

207 Article 62 of the proposed AI Act (n 4); the proposed AI Act (n 4), 36.

208 Article 10(2)(f) and 10(4) of the proposed AI Act (n 4).

209 It has been criticized that although the explanatory memorandum strongly emphasises risks of algorithmic bias, the provisions of the proposed AI Act only moderately refer to performance and publishing of assessments regarding discrimination. MacCarthy and Propp (n 196).

210 The proposed AI Act (n 4), 3.

risk AI systems are stipulated in the proposed AI Act²¹¹. A risk management system would facilitate the identification, analysis and mitigation of risks associated with the AI systems. For instance, if an AI system has the capacity to create discrimination risks, necessary measures can be taken promptly due to the risk management system. In the same vein, providers are required to establish a quality management system for high-risk AI systems²¹². This quality management system involves the systemization and documentation of procedures stipulated in the proposed AI Act²¹³. The quality management system is like a folder that contains all relevant information and documentation regarding the relevant AI system. Therefore, a quality management system may provide a holistic view for the detection and mitigation of AI-related discrimination risks.

4. Obligations for High-risk AI System Users

The proposed AI Act mainly addresses providers of high-risk AI systems in terms of their obligations and their relationship with users²¹⁴. Within this framework, it is also underlined that users must fulfil other obligations under EU laws or national laws²¹⁵. Nevertheless, the proposed AI Act imposes some subsidiary obligations to users explained in the subsequent paragraphs. As seen in the following explanations, these obligations are not sufficiently explained in the proposed AI Act and the performance of these obligations is subject to case-specific instructions given by providers. Because of its unclarity and subsidiarity, this approach may induce a significant uncertainty regarding the compliance responsibilities of users²¹⁶.

According to the proposed AI Act, users²¹⁷ should use the high-risk AI systems in compliance with user's instructions given by providers²¹⁸. Similarly, users are expected to use the information given by providers²¹⁹ for DPIA under Article 35 of the GDPR

211 Article 9 of the proposed AI Act (n 4).

212 Article 17 of the proposed AI Act (n 4).

213 Article 17 of the proposed AI Act (n 4).

214 Chander and Jakubowska (n 195); MacCarthy and Propp (n 196).

215 Article 29(2) of the proposed AI Act (n 4).

216 Dan Whitehead, 'AI & Algorithms (Part 3): Why the EU's AI regulation is a groundbreaking proposal' Hogan Lovells Engage (3 May 2021) <<https://www.engage.hoganlovells.com/knowledgeservices/viewContent.action?key=Ec8tea9VaqLasZjnzVtGsxgHJMKLFEppVpbvVX%2B3OXcP3PYxlq7sZUjdbSm5FletvAtgfIeVU8%3D&nav=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQ0qFfoEM4UR4%3D&emailtofriendview=true&freeviewlink=true>> accessed 20 June 2021.

217 User is defined in Article 3(4) of the proposed AI Act (n 4) as "any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity."

218 Instructions for use are defined in Article 3(15) of the proposed AI Act (n 4) as "the information provided by the provider to inform the user of in particular an AI system's intended purpose and proper use, inclusive of the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used." The proposed AI Act (n 4) specifies the necessary information that should be provided in the user's guide in Article 13(3).

219 Article 13 of the proposed AI Act (n 4).

or Article 27 of Modernised Data Protection Convention 108 as well²²⁰. Users are also required to keep automatically generated log records²²¹. These records may be used as evidence if examination of the AI system's operation procedure or its outcome is required in a legal procedure regarding the investigation of AI-related discrimination claims. Hence, this obligation provides more traceability and transparency for the operation and outcomes of AI-based decision-making procedures.

Furthermore, the proposed AI Act requires users to ensure the relevancy of the input data to the intended purpose of the high-risk AI system, to the extent that users have control over the input data²²². It is an essential obligation from the point of AI-related discrimination because if irrelevant input data is used in an AI system, this system may produce biased decisions. For instance, a person's religion should not be a relevant input in an AI-based decision-making system for employment, except in specific situations²²³. If the religion of a candidate is considered positive or negative input data, this decision-making process produces discriminatory outcomes based on religion. In order to prevent these types of discriminatory risks, users of AI systems are charged with the supervision of the input data's relevancy to the intended purpose of the AI system.

5. Prospective Impacts of the Proposed AI Act on AI-related Discrimination Risks in Private Sector

The list of areas in which the uses of AI systems are qualified as high-risk is provided in Annex III of the proposed AI Act. Education and vocational training, employment, workers management and access to self-employment and access to and enjoyment of essential private services are the areas in the private sector that are listed in Annex III. The Explanatory Memorandum of the proposed AI Act provides explanations regarding the classification of AI systems used in these areas. In the light of these explanations, the impact of the proposed AI Act on AI-related discrimination in the private sector will be discussed in the following paragraphs.

AI systems can be utilized in employment for many purposes such as advertising vacancies, recruitment, and selection, making decisions about promotion, termination and task allocation, assessment of fulfilment of contractual commitments and

²²⁰ Article 29(6) of the proposed AI Act (n 4).

²²¹ The proposed AI Act (n 4), 32.

²²² Article 29(3) of the proposed AI Act (n 4).

²²³ There can be some definite exceptions that a profession has to be performed by a member of a specific religion. For example, an animal slaughtered by a non-Muslim (or nonmember of People of the Book) is not considered halal in mainstream Islamic belief. Therefore, a slaughterhouse may consider candidates' religion in the process for the recruitment of a slaughterer. (The author notes that the ethicality of animal slaughter and meat consumption is getting more debated day by day.) In a similar case, the CJEU ruled that a job application could be rejected only if the nature of the occupational activity concerned or the circumstances in which it is carried out have a strong connection to the ethos of the institution. Also, this requirement must be proportional. See, Case C-414/16, *Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung e.V.* (Judgment of the Court (Grand Chamber) of 17 April 2018).

supervision of behaviours and performance of employees. As it is explained above²²⁴, it is likely that these AI systems may cause discrimination against protected classes. Hence, these systems may risk fundamental rights, including the right to non-discrimination. Meanwhile, these AI systems may affect the professional career and livelihood of persons. Due to these reasons, these AI systems should be classified as high-risk AI systems according to the Explanatory Memorandum of the proposed AI Act²²⁵.

Outcomes of AI systems regarding access to education and vocational training may affect the educational or professional career and thereby, the livelihood of people. As illustrated with the examples above²²⁶, improper design and utilization of these systems may lead to discriminatory outcomes, including reinforcement of previous patterns of discrimination or creating new types of discrimination. Hence, these AI systems may cause a violation of the right to non-discrimination. Therefore, these AI systems should be classified as high-risk systems according to the Explanatory Memorandum of the proposed AI Act²²⁷.

AI systems can be used to determine whether a person can enjoy certain private or public services such as financial resources, accommodation, electricity, and telecommunication services. For instance, a person's credit score or creditworthiness evaluated by AI systems may determine this person's access to these services. As explained above²²⁸, AI-based decision-making in banking may cause discrimination based on sexual orientation, gender, age, and race by either reproducing former patterns of discrimination or producing new types of discrimination. Thus, these AI systems should be qualified as high-risk systems according to the Explanatory Memorandum of the proposed AI Act²²⁹²³⁰. For assessment of creditworthiness and credit scoring, the Explanatory Memorandum of the proposed AI Act sets an exemption for small-scale providers²³¹ who put AI systems into service for their own use. The rationale behind this exemption is the limited degree of their effect and the existence of alternatives on the market²³².

224 See Part IV-1.

225 The proposed AI Act (n 4), 26-27.

226 See Part III-2.

227 The proposed AI Act (n 4), 26.

228 See Part IV-2.

229 The proposed AI Act (n 4), 27.

230 US House Financial Services Committee Chair Maxine Waters cited this approach as an example while she argues that AI discrimination in the lending needs to be tackled with legislation. Ted Knutson, 'AI Lending Discrimination Needs To Be Tackled With Legislation Says House Financial Services Chair' *Forbes* (7 May 2021) <<https://www.forbes.com/sites/tedknutson/2021/05/07/ai-lending-discrimination-needs-to-be-tackled-with-legislation-says-house-financial-services-chair/?sh=7d233f205c7d>> accessed 20 June 2021.

231 The small-scale provider is defined in Article 3(3) of the proposed AI Act (n 4) as a "provider that is a micro or small enterprise within the meaning of Commission Recommendation 2003/361/EC."

232 The proposed AI Act (n 4), 27.

As a result of the above-mentioned classifications of AI systems, in the areas of employment, education, vocational training and finance, as high-risk AI systems, obligations stipulated by the proposed AI act will be imposed to providers and users of these systems. Hence, extra safeguards against AI-related discrimination will be provided for an important part of the AI systems used in the private sector.

6. Other Legal Tools Provided by the Proposed AI Act for the Mitigation AI-related Discrimination Risks

In order to challenge AI-related discrimination, the fact that an AI system is used in the decision-making process should be known in the first place either by the affected individuals or by the relevant authorities. The proposed AI Act provides individuals and the relevant authorities with a right to information about the use of AI systems. Providers of AI systems, regardless of the AI systems' risk level, are required to inform²³³ natural persons regarding their interaction with AI systems, unless it is obvious²³⁴. There are two problems regarding this obligation. First, the nature of the information is not specified. Therefore, providers may confine themselves to a simple notification without detailed information regarding the features of the relevant AI system. Therefore it would be better that the proposed AI Act could specify the content of this disclosure obligation in order to prevent interpretation of this obligation as a simple disclosure notification. Secondly, there should be an “*interaction with AI systems*” to inform the affected persons. Interaction with an AI system is a vague concept. Speaking with chat-bots may be easily thought of as interaction with an AI system. Yet, is it considered interaction with an AI system when an AI system processes personal data to assess persons' creditworthiness or job applications? If the answer is negative, there will be no information requirement toward affected persons regarding the application of many AI systems²³⁵. This concept may be clarified before the legislation of the proposed AI Act. If it does not happen, relevant authorities may provide necessary clarification with their guidelines and decisions. As a last resort, courts would clarify this concept with their interpretations in their decisions.

As an enhancement for the transparency and information obligations, the proposed AI Act ensures accessibility for the relevant authorities, such as market surveillance authorities, national public authorities and bodies, to documentation created under this Act²³⁶. Individuals who learn about their interaction with AI systems and the equality bodies that have access to the documentation respectively may use this information and access authorization to detect and tackle AI-related discrimination risks.

233 The nature of the information is not specified. Therefore, providers may confine themselves with a simple notification without detailed information regarding the features of the relevant AI system.

234 Article 52 of the proposed AI Act (n 4); The proposed AI Act (n 4), 14.

235 MacCarthy and Propp (n 196).

236 Article 64 of the proposed AI Act (n 4), 36.

Although an AI system complies with the proposed AI Act, this AI system may put the protection of fundamental rights in jeopardy. In this case, the proposed AI Act obliges operators of this harmful AI system to take necessary measures to remove this risk²³⁷. As a consequence of this obligation, AI-related discrimination risks are required to be mitigated if these risks occur despite compliance of the AI system with the provisions of the proposed AI Act.

According to the proposed AI Act, the European Commission is entitled to amend specific lists and requirements mentioned in Annexes of the proposed AI Act to enhance adaptation²³⁸. Although this power may adversely affect the predictability regarding the regulation²³⁹, it provides the proposed AI Act with a dynamic nature that is necessary to deal with AI systems. Since the AI field is changing rapidly, emergent AI techniques and approaches may be absorbed into the scope of the proposed AI Act thanks to this opportunity. Hence, any new AI system that may cause a violation of non-discrimination rights can be added to the Annexes and thereby, discrimination risks caused by these emergent AI systems can be mitigated without any delay. In order to enhance the benefit of this dynamic nature in the mitigation of AI-related discrimination, this amending process should include civil society and affected people in a democratic and inclusive way²⁴⁰.

Despite the tools explained in the previous paragraphs, the proposed AI Act has been criticized since it does not introduce a new and more convenient complaint and recourse mechanism for people affected by the AI-based decision-making systems²⁴¹. Moreover, it does not explicitly address the issue of burden of proof in case of alleged breaches²⁴². Since it can be very challenging to prove causal links between the operation of AI-based decision-making systems and the damage, it would be better that the proposed AI Act provides facilitating rules for the benefit of people adversely affected from non-compliance of AI systems to the provisions of the proposed AI Act²⁴³.

VI. Conclusion

In today's market conditions, business organizations need better predictions and faster decision-making processes. AI systems, equipped with ML techniques, may

237 Article 67 of the proposed AI Act (n 4).

238 The proposed AI Act (n 4), 37; Articles 4 and 7 of the proposed AI Act (n 4).

239 Mäkinen (n 7).

240 Chander and Jakubowska (n 195).

241 Chander and Jakubowska (n 195); The European Consumer Organization, 'EU proposal for artificial intelligence law is weak on consumer protection' (Report) (2021) <<https://www.beuc.eu/publications/eu-proposal-artificial-intelligence-law-weak-consumer-protection/html>> accessed 20 June 2021.

242 Yannick Meneceur, 'European Commission's AI regulation proposal: between too much and too little?' *LinkedIn* (23 April 2021) <<https://www.linkedin.com/pulse/european-commissions-ai-regulation-proposal-between-too-meneceur/>> accessed 20 June 2021.

243 For a similar suggestion regarding AI-related discrimination and burden of proof, see Hacker (n 28), 22; Lacroix (n 135), p. 17.

satisfy these needs. However, the integration of AI systems into decision-making processes in the private sector may produce some unintended effects as well. These effects may threaten the protection of some fundamental rights. The right to non-discrimination is one of these fundamental rights at risk. Yet, some fields in the private sector are more vulnerable to discrimination risks caused by the implementation of AI systems. In order to illustrate the significant impact of AI systems on human life and fundamental rights, risky fields in the private sector, namely employment, banking, advertising, pricing and insurance, are investigated with authentic examples of AI-related discrimination in this paper. Hence, it is concluded that the increasing use of AI-based decision-making and the significant influence of these decisions on people's lives and fundamental rights necessitate delicate legal approaches.

Current EU non-discrimination laws and EU data protection laws may provide some tools to mitigate AI-related discrimination risks in the private sector. However, these tools are not enough to tackle specific risks arisen from AI-related discrimination - risks such as intersectional discrimination and proxy discrimination - since these laws are tailored for human discrimination. Although the mechanisms stipulated in these existing laws may be activated through wide interpretations, such an approach can only provide temporary solutions. Therefore, there is an immediate need for new legislation equipped with tools specifically targeting AI-related discrimination risks. In this respect, the proposed AI Act may provide new tools against AI-related discrimination in the private sector since it aims to prevent the breach of fundamental rights via the implementation of AI systems in the public and private sectors. Hence, this paper gives wide coverage to an examination of the proposed AI Act in terms of mitigating AI-related discrimination risks.

According to the risk-based approach adopted by the proposed AI Act, AI systems that may cause a violation of fundamental rights, including the right to non-discrimination, are categorized as high-risk AI systems. Due to the cradle-to-grave approach adopted by the proposed AI Act, providers' and users' of high-risk AI systems are required to comply with specific ex-ante and ex-post obligations. Conformity assessment, CE marking, the establishment of EU-wide public AI registry, human oversight, post-market monitoring, bias monitoring, right-to information about AI-human interactions, right to access for relevant authorities to documentation created under this Act and subsidiary obligations for users, such as compliance with user's instructions given by providers, supervision of automatically generated log records and ensuring relevancy of the input data to the intended purpose of AI system are some of these obligations.

In this paper, the said obligations are analysed in terms of their prospective impacts on the mitigation of AI-related discrimination risks. It was found that provisions of the proposed AI Act can provide new legal safeguards against AI-related discrimination risks in the private sector. However, these legal safeguards are not adequate in the

face of intractable AI-related discrimination risks. Therefore, the vital need for new legislation regarding mitigation of AI-related discrimination risks is still present. It is expected that criticism regarding AI-related discrimination risks, indicated in this paper and other papers, should be considered in the legislation process of the proposed AI Act and other new legislations.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The author has no conflict of interest to declare.

Financial Disclosure: The author declared that this study has received no financial support.

Acknowledgement: This article is an extended and revised version of a paper prepared for Recht der Digitalisierung seminar at the Philipps-University of Marburg, Germany. The author sincerely thanks Prof. Dr. Florian Möslin, LL.M. (London) and Prof. Dr. Johannes Buchheim, LL.M. (Yale), for their insightful feedback and encouraging comments.

Bibliography

- Accesswire, '72% of Resumes are Never Seen by Employers' (16 February 2016) <<https://www.accesswire.com/436847/72-of-Resumes-are-Never-Seen-by-Employers>> accessed 20 June 2021.
- Ali M et al., 'Discrimination Through Optimization: How Facebook's ad Delivery Can Lead to Skewed Outcomes' (2019) Proceedings of the ACM on Human-Computer Interaction. <<https://arxiv.org/pdf/1904.02095.pdf>> accessed 20 June 2021.
- Angelopoulos C et al., 'Study Of Fundamental Rights Limitations For Online Enforcement Through Self-Regulation' (Study Report of Institute for Information Law, University of Amsterdam) (2016) <<https://scholarlypublications.universiteitleiden.nl/access/item%3A2869513/view>> accessed 20 June 2021.
- Angwin J, Mattu S and Larson J, 'The tiger mom tax: Asians are nearly twice as likely to get a higher price from Princeton Review'. *ProPublica* (1 September 2015) <<https://www.ProPublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review>> accessed 20 June 2021.
- Angwin J, Scheiber N and Tobin A, 'Dozens of companies are using Facebook to exclude older workers from job ads' *ProPublica* (20 December 2017) <<https://www.propublica.org/article/facebook-ads-age-discrimination-targeting>> accessed 20 June 2021.
- Angwin J, Tobin A and Varner M, 'Facebook (still) letting housing advertisers exclude users by race' *ProPublica* (21 November 2017) <<https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>> accessed 20 June 2021.
- Avraham R, 'Discrimination and Insurance' in Lippert-Rasmussen K. (ed.), *The Routledge Handbook of the Ethics of Discrimination*, (Routledge, UK 2018), pp. 335-347.
- Bartlett R et al., 'Consumer-Lending Discrimination in the FinTech Era' (2019) (National Bureau of Economic Research Working Papers 25943).
- Barocas S and Selbst A D, 'Big Data's Disparate Impact' (2016) 104 California Law Review, 671-732.
- Bathae Y, 'The Artificial Intelligence Black Box And The Failure Of Intent And Causation' (2018) 31:2 Harvard Journal of Law & Technology, 890-938.
- Beck S et al., 'Künstliche Intelligenz und Diskriminierung Herausforderungen und Lösungsansätze' (2019) (Whitepaper from Plattform Lernende Systeme) <<https://www.plattform-lernende-systeme.de/publikationen-details/kuenstliche-intelligenz-und-diskriminierung-herausforderungen-und-loesungsansaezte.html>> accessed 20 June 2021.

- Biao v. Denmark (Grand Chamber)* (App no 38590/10) ECHR 24 May 2016.
- Bodo B et al., 'Tackling the algorithmic control crisis-the technical, legal, and ethical challenges of research into algorithmic agents' (2017) 19 *Yale Journal of Law & Technology*, 133-180.
- Borgesius F Z, 'Discrimination, Artificial Intelligence and Algorithmic Decision-making' (Study Report) (Council of Europe, Strasbourg 2018) <<https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>> accessed 20 June 2021.
- Borgesius F J Z, 'Strengthening legal protection against discrimination by algorithms and artificial intelligence' (2020) 24 10 *The International Journal of Human Rights*, 1572- 1593.
- Bundesanstalt für Finanzdienstleistungsaufsicht, 'Big data meets artificial intelligence: Challenges and implications for the supervision and regulation of financial services' (2018) (Study) <https://www.bafin.de/SharedDocs/Downloads/EN/dl_bdai_studie_en.html> accessed 20 June 2021.
- Burrell J, 'How the machine 'thinks': understanding opacity in machine learning algorithms' (2016) *Big Data & Society* Doi:10.1177/2053951715622512.
- Case 170/84, *Bilka-Kaufhaus GmbH and Karin Weber von Hartz*, (Judgment of the Court of 13 May 1986.)
- Case C-83/14, "*CHEZ Razpredelenie Bulgaria*" *AD v Komisia za zashtita ot diskriminatsia*, (Judgment of the Court (Grand Chamber) of 16 July 2015).
- Case C443/15, *David L. Parris v Trinity College Dublin and Others* (Judgment of the Court (First Chamber) of 24 November 2016).
- Case C-414/16, *Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung e.V.* (Judgment of the Court (Grand Chamber) of 17 April 2018).
- Chander S and Jakubowska E, 'EU's AI law needs major changes to prevent discrimination and mass surveillance' *EDRI* (2021) <<https://edri.org/our-work/eus-ai-law-needs-major-changes-to-prevent-discrimination-and-mass-surveillance/>> accessed 20 June 2021.
- Charter of Fundamental Rights of the European Union, 2012/C 326/02 OJ C 326, [2012].
- Commission (EC) 'Proposal for a Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts' COM(2021) 206 final, 21 April 2021.
- Corner N, 'Revealed: How your name, your ethnicity, and even your e-mail address could be adding hundreds of pounds to your car insurance premium' *Dailymail* (29 June 2018) <<https://www.dailymail.co.uk/femail/article-5896561/How-email-address-add-hundreds-car-insurance-premium.html>> accessed 20 June 2021.
- Council Directive (EC) 2000/43 of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L 180.
- Council Directive (EC) 2000/78 of 27 November 2000 establishing a general framework for equal treatment in employment and occupation [2000] OJ L 303.
- Council Directive (EC) 2001/95 of the European Parliament and of the Council of 3 December 2001 on general product safety [2001] OJ L 11.
- Council of Europe, 'Explanatory Report to the Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms' (2000) <<https://rm.coe.int/09000016800cce48>> accessed 20 June 2021.
- Council of Europe, 'Algorithms and Human Rights: Study On The Human Rights Dimensions Of Automated Data Processing Techniques (In Particular Algorithms) And Possible Regulatory Implications' (Study Report) (DGI(2017)12) (2018).

- Council Regulation (EC) 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.
- Craglia M et al., ‘Artificial Intelligence: A European Perspective’ (Joint Research Centre Working Papers) (2018).
- Danks D and London A J, ‘Algorithmic Bias in Autonomous Systems’ (2017) IJCAI’17: Proceedings of the 26th International Joint Conference on Artificial Intelligence, 4691-4697.
- Dastin J, ‘Amazon scraps secret AI recruiting tool that showed bias against women’ *Reuters* (11 October 2018) <<https://www.reuters.com/article/us-amazon-com-jobs-automation-in...-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>> accessed 20 June 2021.
- Data Protection Working Party Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 Wp251rev.01. [2018].
- Datta A, Tschantz M C and Datta A, ‘Automated Experiments on Ad Privacy Settings A Tale of Opacity, Choice, and Discrimination’ (2015) Proceedings on Privacy Enhancing Technologies <<https://www.andrew.cmu.edu/user/danupam/dtd-pets15.pdf>> accessed 20 June 2021.
- Delipetrev B, Tsinarakii C and Kostić U, ‘Historical Evolution of Artificial Intelligence’ (Technical Report) (Publications Office of the European Union, Luxembourg 2020) <https://publications.jrc.ec.europa.eu/repository/handle/JRC120469> accessed 20 June 2021.
- Dwork C et al., ‘Fairness Through Awareness, Proceedings of the 3rd Innovations in Theoretical Computer Science Conference’ (2012) ITCS ‘12 (Association for Computing Machinery), 214–226.
- Edwards, L and Veale M, ‘Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for’ (2017) 16 *Duke Law & Technology Review*, 18-84.
- Espinoza J and Murgia M, ‘Europe attempts to take leading role in regulating uses of AI’ *Financial Times* (24 April 2021) <<https://www.ft.com/content/360faa3e-4110-4f38-b618-dd695dece90>> accessed 20 June 2021.
- European Commission, Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation [2008] {SEC(2008) 2180} {SEC(2008) 2181} /* COM/2008/0426 final - CNS 2008/0140 */.
- European Commission, ‘Special Eurobarometer Discrimination in the EU in 2015’ (Report) (2015).
- European Commission’s High-Level Expert Group on Artificial Intelligence, ‘A Definition of AI: Main Capabilities and Disciplines’ (2019) (Report) <<https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>> accessed 20 June 2021.
- European Commission, ‘A Union of Equality: Gender Equality Strategy 2020–2025’ [2020] COM (2020) 152.
- European Commission, ‘Internal Market, Industry, Entrepreneurship and SMEs’ <https://ec.europa.eu/growth/single-market/ce-marking/manufacturers_en> accessed 20 June 2021.
- European Commission, ‘Harmonised Standards’ <https://ec.europa.eu/growth/single-market/european-standards/harmonised-standards_en> accessed 20 June 2021.
- European Convention for the Protection of Human Rights and Fundamental Freedoms (1950) <https://www.echr.coe.int/Documents/Convention_ENG.pdf> accessed 20 June 2021.

- European Union Agency for Fundamental Rights, '#BigData: Discrimination in data-supported decision making' (Report) (2018) Doi:10.2811/343905.
- European Union Agency for Fundamental Rights, 'Getting the Future Right-Artificial Intelligence and Fundamental Rights' (2020) Doi:10.2811/774118.
- Fairness, Accountability, and Transparency in Machine Learning <<https://www.fatml.org/>> accessed 20 June 2021.
- Favaretto M, De Clercq E and Elger B S, 'Big Data and discrimination: perils, promises and solutions. A systematic review' (2019) 6 *Journal of Big Data* <<https://doi.org/10.1186/s40537-019-0177-4>> accessed 20 June 2021.
- Financial Conduct Authority, 'Pricing practices in the retail general insurance sector: Household insurance' (2018) <<https://www.fca.org.uk/publication/thematic-reviews/tr18-4.pdf>> accessed 20 June 2021.
- Stan F, 'History, Motivations, and Core Themes' in Keith Frankish and William W. Ramsey, (eds.), *The Cambridge Handbook Of Artificial Intelligence* (Cambridge University Press, UK 2014).
- Fuster A et al., 'Predictably Unequal? The Effects of Machine Learning on Credit Markets' (2020). *SSRN* <<https://ssrn.com/abstract=3072038>> accessed 20 June 2021.
- Future of Life Institute, 'The Asilomar AI principles' (2017) <<https://futureoflife.org/ai-principles/>> accessed 20 June 2021.
- Gillis T B and Spiess J L, 'Big Data and Discrimination' (2019) 86 *The University of Chicago Law Review*, pp. 459-487.
- Hacker P., 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law' (Pre-print version) (2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3164973&download=yes> accessed 20 June 2021.
- Heinrichs B, 'Discrimination in the age of artificial intelligence' (2021) *AI & Society* Springer. <<https://doi.org/10.1007/s00146-021-01192-2>> accessed 20 June 2021.
- Hildebrandt M and Gutwirth S, 'Concise Conclusions: Citizens Out of Control' in Hildebrandt M and Gutwirth S (eds.) *Profiling the European Citizen: Cross-Disciplinary Perspectives* (Springer, 2008).
- Hill K, 'How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did' *Forbes* (16 February 2012) <<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=415c8bfe6668>> accessed 20 June 2021.
- Humphreys P, 'The philosophical novelty of computer simulation methods' (2008) 169 *Synthese*, pp. 615-626 <<https://doi.org/10.1007/s11229-008-9435-2>> accessed 20 June 2021.
- International Association of Insurance Supervisors, 'Issues Paper on the Use of Big Data Analytics in Insurance' (2020) <<https://www.iaisweb.org/page/supervisory-material/issues-papers//file/89244/issues-paper-on-use-of-big-data-analytics-in-insurance>> accessed 20 June 2021.
- International Covenant on Civil and Political Rights adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of 16 December 1966 entry into force 23 March 1976.
- Izzidien R, 'Higher insurance if you're called Mohammed? That's just the start of institutionalized Islamophobia' *NewStatesman* (23 January 2018) <<https://www.newstatesman.com/politics/uk/2018/01/higher-insurance-if-you-re-called-mohammed-s-just-start-institutionalised>> accessed 20 June 2021.
- Johnson K, Pasquale F and Chapman J, 'Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation' (2019) 88 2 *Fordham Law Review* 499-529.

- Jones H, 'When the wrong postcode puts insurance out of your reach' *The Guardian* (30 October 2016) <<https://www.theguardian.com/money/2016/oct/30/wrong-postcode-puts-insurance-out-of-reach>> accessed 20 June 2021.
- Kelion L, 'Apple's 'sexist' credit card investigated by US regulator' *BBC* (11 November 2019) <<https://www.bbc.co.uk/news/business-50365609>> accessed 20 June 2021.
- Kim P T, 'Data-Driven Discrimination at Work' (2017) 58 3 *William & Mary Law Review*, pp. 857-936.
- Kleinberg J et al., 'Discrimination in the Age of Algorithms' (2018) 10 *Journal of Legal Analysis* pp. 113-174.
- Kloza D et al., 'Data protection impact assessments in the European Union: complementing the new legal framework towards a more robust protection of individuals' (d.pia.lab Policy Brief) (2017) <https://cris.vub.be/files/32009890/dpialab_pb2017_1_final.pdf> accessed 20 June 2021.
- Knutson T, 'AI Lending Discrimination Needs To Be Tackled With Legislation Says House Financial Services Chair' *Forbes* (7 May 2021) <<https://www.forbes.com/sites/tedknutson/2021/05/07/ai-lending-discrimination-needs-to-be-tackled-with-legislation-says-house-financial-services-chair/?sh=7d233f205c7d>> accessed 20 June 2021.
- Kretchmer H, 'Insurers 'risk breaking racism laws'' *BBC* (9 February 2018) <<https://www.bbc.com/news/business-43011882>> accessed 20 June 2021.
- Kroll J A et al., 'Accountable Algorithms' (2017) 165 *University of Pennsylvania Law Review*, 633-705.
- Lacroix C, 'Preventing discrimination caused by the use of artificial intelligence' (Report of Committee on Equality and Non-Discrimination at Council of Europe) (2020).
- Lambrecht A and Tucker C E, 'Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads' (2019) 65 7 *Management Science*, 2966-2981.
- Liboreiro J, 'The higher the risk, the stricter the rule': Brussels' new draft rules on artificial intelligence' *Euronews* (21.4.2021) <<https://www.euronews.com/2021/04/21/the-higher-the-risk-the-stricter-the-rule-brussels-new-draft-rules-on-artificial-intelligence>> accessed 20 June 2021.
- Loi M and Christen M, 'Choosing how to discriminate: navigating ethical trade-offs in fair algorithmic design for the insurance sector' (2021) *Philosophy & Technology* <<https://doi.org/10.1007/s13347-021-00444-9>> accessed 20 June 2021.
- Lucci S and Kopec D, *Artificial Intelligence in the 21st Century: A Living Introduction* (Mercury Learning and Information, USA 2016).
- MacCarthy M and Propp K, 'Machines Learn That Brussels Writes the Rules: The EU's New AI Regulation' *Lawfare Blog* (28 April 2021) <<https://www.lawfareblog.com/machines-learn-brussels-writes-rules-eus-new-ai-regulation>> accessed 20 June 2021.
- Mäkinen J, 'The EU's AI Regulation Proposal – How to Fix It?' (2021) <<https://teknologiateollisuus.fi/en/ajankohtaista/eus-ai-regulation-proposal-how-fix-it>> accessed 20 June 2021.
- Meneceur Y, 'European Commission's AI regulation proposal: between too much and too little?' *LinkedIn* (23 April 2021) <<https://www.linkedin.com/pulse/european-commissions-ai-regulation-proposal-between-too-meneceur/>> accessed 20 June 2021.
- Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data CM/Inf(2018)15-final [2018].
- Möslein F, 'Robots in the Boardroom: Artificial Intelligence and Corporate Law' in Barfield W and Pagallo U (eds.), *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar, UK 2018) 649-669.

- Newell S and Marabelli M, 'Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'' (2015) 24 *Journal of Strategic Information Systems*, 3-15.
- Peck D, 'They're Watching You at Work' *Atlantic* (December 2013) <<http://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681>> accessed 20 June 2021.
- Peppet S R, 'Regulating the Internet of Things: first steps toward managing discrimination, privacy, security, and consent' (2014) 93 *Texas Law Review*, 85-176.
- Pettinger-Kearney N et al., 'The EU's proposed AI Regulation' (2021) <<https://www.freshfields.com/en-gb/our-thinking/campaigns/digital/artificial-intelligence/the-eus-proposed-ai-regulation/>> accessed 20 June 2021.
- Pope D G and Sydnor J R, 'Implementing Anti-Discrimination Policies in Statistical Profiling Models' (2011) 3 *American Economic Journal: Economic Policy*, 206–231.
- Prince A E R and Schwarcz D, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105 *Iowa Law Review*, 1257-1318.
- Probasco J, 'The Insurance Industry Confronts Its Own Racism' *Investopedia* (1 September 2020) <<https://www.investopedia.com/race-and-insurance-5075141#citation-9>> accessed 20 June 2021.
- Reinhold F and Müller A, 'Algorithm Watch's response to the European Commission's proposed regulation on Artificial Intelligence – a major step with major gaps' (Report) (22 April 2021) <<https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021/>> accessed 20 June 2021.
- Rieke A, Bogen M and Robinson D G, 'Public scrutiny of automated decisions: Early lessons and emerging methods' (Upturn and Omidyar Network Report) (2018) <<https://www.data.govt.nz/assets/Uploads/Public-Scrutiny-of-Automated-Decisions.pdf>> accessed 20 June 2021.
- Rodrigues R, 'Legal and human rights issues of AI: Gaps, challenges and vulnerabilities' (2020) 4 *Journal of Responsible Technology* <<https://doi.org/10.1016/j.jrt.2020.100005>> accessed 20 June 2021.
- Samoili S et al., 'AI Watch. Defining Artificial Intelligence. Towards An Operational Definition And Taxonomy Of Artificial Intelligence' (Joint Research Centre Technical Reports) (Luxembourg 2020).
- Schauer F, 'Statistical (and non-statistical) discrimination' in Kasper Lippert-Rasmussen (ed.), *The Routledge Handbook of the Ethics of Discrimination*, (Routledge, UK 2018), 42-53.
- Swedloff R, 'Risk classification's Big Data (r)evolution' (2014) 21 *Connecticut Insurance Law Journal*, 339-373.
- The European Consumer Organization, 'Automated Decision Making And Artificial Intelligence - A Consumer Perspective' (Report) (2018) <https://www.beuc.eu/publications/beuc-x-2018-058_automated_decision_making_and_artificial_intelligence.pdf> accessed 20 June 2021.
- The European Consumer Organization, 'AI Rights for Consumers' (Report) (2019) <https://www.beuc.eu/publications/beuc-x-2019-063_ai_rights_for_consumers.pdf> accessed 20 June 2021.
- The European Consumer Organization, 'The Use Of Big Data And Artificial Intelligence In Insurance' (Report) (2020) <https://www.beuc.eu/publications/beuc-x-2020-039_beuc_position_paper_big_data_and_ai_in_insurances.pdf> accessed 20 June 2021.
- The European Consumer Organization, 'EU proposal for artificial intelligence law is weak on consumer protection' (Report) (2021) <<https://www.beuc.eu/publications/eu-proposal-artificial-intelligence-law-weak-consumer-protection/html>> accessed 20 June 2021.

- The Federal Government of Germany, 'Artificial Intelligence Strategy' (2018) <https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf> accessed 20 June 2021.
- The Montreal Declaration for Responsible AI (2017) <<https://www.montrealdeclaration-responsibleai.com/the-declaration>> accessed 20 June 2021.
- The Partnership on AI to Benefit People and Society <<https://www.partnershiponai.org/about/>> accessed 20 June 2021.
- The Royal Society, 'Machine Learning: The Power And Promise Of Computers That Learn By Example' (Report) (April 2017) <royalsociety.org/machine-learning> accessed 20 June 2021.
- The Royal Society, 'Explainable AI: the basics' (Policy briefing) (2019) <royalsociety.org/ai-interpretability/> accessed 20 June 2021.
- The White House, 'Big Data: Seizing Opportunities, Preserving Values' (Report) (2014).
- Universal Declaration of Human Rights, UNGA Res 217/A (10 December 1948).
- Valentino-Devries J, Singer-Vine J, and Soltani A, 'Websites vary prices, deals based on users' information' *Wall Street Journal* (24 December 2012) <<https://www.wsj.com/articles/SB10001424127887323777204578189391813881534>> accessed June 2021.
- Vigdor N, 'Apple Card Investigated After Gender Discrimination Complaints' *The New York Times* (10 November 2019) <<https://www.nytimes.com/2019/11/10/business/apple-credit-card-investigation.html>> accessed June May 2021.
- Wachter S, Mittelstadt B and Russell C, 'Counterfactual explanations without opening the black box: automated decisions and the GDPR' (2018) 31 2 *Harvard Journal of Law & Technology*, 842-887.
- Wachter S, Mittelstadt B and Russell C, 'Why Fairness Cannot Be Automated: Bridging The Gap Between EU Non-Discrimination Law And AI' (Draft) (2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3547922> accessed 20 June 2021.
- Wagner B et al., 'Algorithms and human rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications' (Study Report) (DGI(2017)12) (2018) <<https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>> accessed 20 June 2021.
- Wahlster W and Winterhalter C, 'German Standardization Roadmap On Artificial Intelligence' (German Institute for Standardization and DKE German Commission for Electrical, Electronic & Information Technologies of DIN and VDE November 2020) <<https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsmroadmap-en-data.pdf>> accessed 20 June 2021.
- Walch K, 'Is There A Difference Between Assisted Intelligence Vs. Augmented Intelligence?' *Forbes*. (12 January 2020) <<https://www.forbes.com/sites/cognitiveworld/2020/01/12/is-there-a-difference-between-assisted-intelligence-vs-augmented-intelligence/?sh=3300950426ab>> accessed 20 June 2021.
- Wang P, 'On the Working Definition of Intelligence' (1995) <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.5012>> accessed 20 May 2021.
- Weisbaum H, 'Data Mining Is Now Used to Set Insurance Rates; Critics Cry Foul' *CNBC* (16 April 2014) <<https://www.cnn.com/2014/04/16/data-mining-is-now-used-to-set-insurance-rates-critics-cry-fowl.html#:~:text=You%20might%20not%20like%20the,to%20provide%20you%20that%20coverage>> accessed 20 June 2021.

Whitehead D, 'AI & Algorithms (Part 3): Why the EU's AI regulation is a ground breaking proposal' *Hogan Lovells Engage* (3 May 2021) <<https://www.engage.hoganlovells.com/knowledgeservices/viewContent.action?key=Ec8teaJ9VaqLasZjnzVtGsxgHJMKLFEppVpbbVX%2B3OXcP3PYxlq7sZUjdb-Sm5FIetvAtgf1eVU8%3D&nav=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQ0qF-foEM4UR4%3D&emailtofriendview=true&freeviewlink=true>> accessed 20 June 2021.

Williams B et al., 'How Algorithms Discriminate Based on Data they Lack: Challenges, Solutions, and Policy Implications' (2018) 8 *Journal of Information Policy*, 78-115.

World Economic Forum, 'Transforming Paradigms A Global AI in Financial Services Survey' (2020) (Report) <http://www3.weforum.org/docs/WEF_AI_in_Financial_Services_Survey.pdf> accessed 20 June 2021.

Xenidis R, 'Tuning EU equality law to algorithmic discrimination: Three pathways to resilience' (2020) 27(6) *Maastricht Journal of European and Comparative Law*, 736-758.

Xenidis R and Senden L, 'EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination' in Ulf Bernitz et al. (eds.), *General principles of EU law and the EU digital order* (Kluwer Law International 2020), 151-182.