Automated Decision-Making in Online Platforms: Protection Against Discrimination and Manipulation of Behavior

19685 words

Table of Contents

Executive Summary

As technology advanced, people increasingly relied on the Internet daily. Online platforms such as search engines, e-commerce websites, social media, and on-demand services have become integral to millions of people's lives. These platforms utilize algorithms and machine learning to offer personalized experiences to users through automated decision-making (ADM). However, despite their effective nature, there have been problematic discrimination and behavior manipulation associated with these ADM.

This thesis focuses on the EU approach to addressing discrimination and manipulation of behavior arising from using ADM in online platforms. The main research question of this thesis is:

"To what extent is the current EU legal framework for data protection and the proposed artificial intelligence regulation adequate to address discrimination and manipulation of behavior arising from automated decision-making (ADM) used in online platforms?".

To answer this question, the thesis relies on desk research. It primarily analyzes the General Data Protection Regulation (GDPR) and the proposed Artificial Intelligence Act (AIA) to address the sufficiency of the regulations against discrimination and manipulation of behavior arising from ADM on online platforms.

The thesis reveals that ADM is a system that automates decisions about individuals through underlying technologies such as rule-based or machine learning algorithms. Although ADM provides efficient outcomes, it risks carrying biases, generating inaccurate outcomes, and inferring data about individuals that might lead to behavior manipulation and discrimination. This thesis distinguishes problematic behavior manipulation practices on online platforms and finds that there are morally acceptable manipulative practices, including personalized recommendations, among problematic ones, such as personalized advertisements that exploit individuals' vulnerabilities. For discriminatory practices, it highlights two problematic areas: discrimination based on *protected characteristics,* such as race, and discrimination based on *non-protected-characteristic*s, such as socioeconomic status.

The thesis examines the GDPR and the AIA and explores how discrimination and manipulation of behavior arising from using ADM are regulated. The GDPR regulates ADM with a prohibition and empowers individuals with the right to information and access concerning ADM. The AIA, in contrast, focuses on the underlying technologies and regulates their impact on individuals. This thesis finds there is no need for a new regulation to address discrimination and manipulation of behavior arising from ADM used in online platforms. However, it provides recommendations to the GDPR and the AIA for clear and comprehensive rules. For the GDPR, this thesis recommends clarity for the specific rules for ADM under Article 22(1) GDPR and the addition of minimum safeguards to Article 22(3) GDPR to increase protection. For Articles 13,14, and 15 GDPR, clarity over concepts used and a clear distinction between *ex-ante* and *ex-post* information to include user-centric transparency is suggested. For the AIA, there are five recommendations, which include a clear AI system definition that acknowledges the underlying technologies, a definition for recommender systems, the addition of very-large-online-search-engines for a complete online platform representation, clarity over significant

harm requirement for prohibited AI systems and clear grounds on high-risk AI system requirements to balance technical feasibilities of the underlying technologies. The thesis also underlines the importance of increased enforcement of existing GDPR rules, such as general principles and data protection impact assessments, and the need to thoroughly interpret both regulations for a beneficial AI system application.

In conclusion, this thesis emphasizes the need for improvements to the GDPR and the AIA, additionally the emerging need for comprehensive interpretations of both regulations for unified enforcement over automated decision-making systems in online platforms.

Main Findings

- Automated decision-making ("**ADM**") is a system that uses rule-based or machine learning algorithms that generate decisions about individuals. Most online platforms opt for a machine learning algorithm for the ADM they have in place.
- ADM can infer data about individuals and create different outcomes due to their learning capabilities. Embedded biases could result in errors, not only in algorithm design but also in different ADM stages. ADM with machine learning capabilities might be inexplicable, creating a "black box effect" that hinders the explainability and transparency of the ADM.
- Manipulative practices on online platforms are not always problematic; some can be considered morally acceptable. In contrast, discriminatory practices, regardless of the discriminatory ground being *protected characteristics* or not, shall always be an unjust treatment of an individual.
- Due to the underlying technology of the ADM, there might be unintentional influences alongside intentional ones. Thus, ADM on online platforms can intentionally manipulate and discriminate against people while risking unintentional influences.
- The General Data Protection Regulation ("**GDPR**") explicitly regulates ADM but does not regulate the impact of ADM on individuals; however, it empowers individuals through data subject rights. The Proposed Artificial Intelligence Act ("**AIA**") does not regulate ADM directly but focuses on the risks and impacts of the underlying technologies, including discriminatory and manipulative AI systems.
- Although the GDPR regulates ADM in a general manner, the wording creates ambiguity. The GDPR also provides two different levels of protection to individuals; however, the similar wording creates interpretation issues for the application of the DSRs. In contrast, the AIA regulates discriminatory and manipulative AI practices with a prohibition, as well as recommender systems of online platforms as high-risk AI systems, although there is a vague approach to these articles that needs clarity on several aspects.

1. Introduction

This chapter provides background information on automated decision-making systems by providing a general understanding, contextualizing automated decision-making systems used in online platforms, describing problems raised by such use in Section 1.1., formulates the research objective and central research question and sub-questions in Section 1.2. and 1.3., respectively. This chapter also clarifies the methodology used in Section 1.4. and introduces structure of the thesis in Section 1.5.

1.1. Setting the stage

Artificial intelligence ("**AI**") has been a topic of exploration for several decades, with its complexity and applications still being researched today. The concept of AI is familiar, as ancient civilizations like Ancient Greece contemplated the existence of intelligent robots and artificial beings.[1] As time went by and electronic computing began to develop, Alan Turing, assigned to crack the ENIGMA code during World War II, began to consider the question, "Can machines think?" back in 1950."[2] This marked the beginning of AI as it is understood today. IBM's Deep Blue,[3] or movies about robots, shape popular ideas on AI; however, its complexity goes beyond robots. AI is currently used in diverse applications through algorithms capable of self-learning, which operate without our awareness. AI adoption in businesses doubled compared to 2017, reaching a 2.5-time increase.[4] Companies are automating various data-related processes for efficiency and generating high-quality data for their algorithms.[5] This indicates a growing focus on 'automated decision-making' ("**ADM**") tools.

ADM is a system that generates automated decisions without direct human involvement.[6] It relies on factual data, training datasets, or inferred data.[7] Typically, an ADM uses algorithms, rules, or data to produce an outcome with pre-defined rules.[8] Depending on the algorithms, ADM can additionally have machine learning capabilities that enable the system to learn from its environment without any explicit programming[9] enhancing its predictive abilities and potentially generating more accurate or new outcomes.[10] ADM often has profiling capabilities

---

[1] 'Antiquity' (*A History of Artificial Intelligence*) <https://ahistoryofai.com/antiquity/> accessed 7 March 2023.

[2] Alan M Turing, 'Computing Machinery and Intelligence' [1950] Computing Machinery and Intelligence 433.

[3] 'IBM100 - Deep Blue' <https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/> accessed 7 March 2023.

[4] 'The State of AI in 2022 - and a Half Decade in Review' <http://ceros.mckinsey.com/mckinsey-commentary-ai-meyhew-mobile-1-1> accessed 7 March 2023.

[5] ibid.

[6] 'What Is Automated Individual Decision-Making and Profiling?' (2022) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/> accessed 7 March 2023.

[7] ibid.

[8] ibid.

[9] 'AI & Machine Learning' (*Center for Democracy and Technology*) <https://cdt.org/ai-machine-learning/> accessed 7 March 2023.

[10] 'What Is Machine Learning and Why Is It Important?' (*Enterprise AI*) <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML> accessed 7 March 2023.

that involve analyzing data on personality, behavior, and interests to predict and generate decisions about individuals.[11]

As AI evolves and gains worldwide use, society increasingly interacts with them primarily through social media, search engines, and e-commerce websites ("**online platforms**"). Although this interaction positively impacts individuals with personalized content, it can also negatively impact them.

Manipulation of behavior involves techniques to guide individuals away from their rational decision-making. ADM used in recommender systems of online platforms for advertisement purposes uses data and generates relevant personalized and targeted advertisements to manipulative individuals. Manipulation of behavior has been extensively discussed in recent years as individuals increasingly encounter digital "behavioral" technologies that aim to change or support change in human behaviors.[12] Online platforms create dynamic, interactive, and personal mediums that facilitate the exploitation of individual vulnerabilities through ADM.[13] Manipulation of behavior on online platforms can manifest as content moderation, where the system influences what individuals see on their "personalized" feeds. While personalized recommendations from Spotify and Netflix can be welcomed for their convenience,[14] other manipulation tactics through ADM can neither expected nor considered fair. The Facebook-Cambridge Analytica scandal is a notable example of such problematic practices.

ADM with embedded bias, on the other hand, whether through biased training datasets or technical errors,[15] can lead to discriminatory outcomes rather than just being manipulative. Examples include Google Photos categorizing African-Americans as "apes"[16] or Facebook targeting job advertisements that discriminate against individuals based on age, nationality, and gender.[17] Despite non-discrimination being prohibited under many international, regional, and national legislations, ADM can still produce discriminatory practices outside these

---

[11] Jason Brownlee, 'Difference Between Algorithm and Model in Machine Learning' (*MachineLearningMastery.com*, 2020) <https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/> accessed 7 March 2023.

[12] Michael Klenk, '(Online) Manipulation: Sometimes Hidden, Always Careless' (2022) 80 Review of Social Economy 85 <https://www.tandfonline.com/doi/full/10.1080/00346764.2021.1894350> accessed 7 March 2023.

[13] ibid.

[14] Fleur Jongepier, Micheal Klenk, 'The Philosophy of Online Manipulation', Fleur Jongepier, Micheal Klenk (eds), *The Philosophy of Online Manipulation* (1st edn, Routledge, 2022) <https://www.taylorfrancis.com/books/9781003205425> accessed 19 March 2023.

[15] B. Goodman, Discrimination, Data Sanitisation and Auditing in the European Union's General Data Protection Regulation, European Data Protection Law Review, 495, 2nd Volume (2016) < https://edpl.lexxion.eu/article/edpl/2016/4/8> accessed 7 March 2023.

[16] Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York University Press 2018).

[17] 'Are Facebook Job Ads Discriminatory? Company Accused of Bias against Women, Older Workers' (*USA TODAY*) <https://www.usatoday.com/story/money/2022/12/01/facebook-jobs-ads-discrimination-women-older-workers/10810589002/> accessed 22 March 2023.

prohibitions. This can be seen in discriminatory housing ads on online platforms that target individuals profiled as "living in poverty,"[18] which is inherently unfair but not prohibited.

The invisibility and inexplicability of ADM aggravate discriminatory or manipulative practices. These raise questions about whether decisions derived from ADM are accurate, lawful, fair, and unbiased toward the individuals in their respective contexts. Individuals' concerns over excessive data collection by big tech companies increased their trust in data protection regulations in the EU. This made individuals pursue answers within the well-known legislation that empowers individuals.[19] Because of this societal trust level and the empowerment of individuals, this thesis primarily analyzes data protection legislation. Another pillar to the basis of this thesis is the recognition of the significance of AI integration and the anticipated AI legislation[20] that includes prohibitions on certain AI applications. While there may be other legislations that would help mitigate such risks, such as non-discrimination laws, they have limitations such as the narrow scope of *protected characteristics*, new forms of discrimination, and the need for tailored articles addressing ADM-specific concerns. However, when it comes to behavior manipulation, the lack of specific legislation, like non-discrimination laws, creates the necessity to use data protection regulation and proposed AI legislation to establish a framework.

Due to the societal trust in data protection regulations and awaited AI legislation, this thesis focuses on analyzing *General Data Protection Regulation*[21] ("**GDPR**") and the proposed *Artificial Intelligence Act*[22] ("**AIA**") to examine how discrimination and manipulation of behavior arising from the use of ADM in online platforms is addressed. The analysis also includes further steps to address discrimination and manipulation of behavior in cases where the regulations might not be adequate to address such problems.

1.2. Research objective

This thesis analyzes whether the EU legal framework can address the discriminatory and manipulative outcomes of ADM used in online platforms. This thesis also investigates whether

---

[18] 'Poverty Inextricably Linked to Discrimination and Racism – UN Special Rapporteur' (*OHCHR*) <https://www.ohchr.org/en/press-releases/2013/11/poverty-inextricably-linked-discrimination-and-racism-un-special-rapporteur> accessed 22 March 2023.

[19] M Hawath, 'Regulating Automated Decision-Making: An Analysis of Control over Processing and Additional Safeguards in Article 22 of the GDPR.' (2021) 7 European Data Protection Law Review 161 <https://edpl.lexxion.eu/article/EDPL/2021/2/6> accessed 20 March 2023.

[20] 'A European Approach to Artificial Intelligence | Shaping Europe's Digital Future' (2023) <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> accessed 10 May 2023.

[21] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679&qid=1688059883993> accessed 10 March 2023.

[22] Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) And amending certain Union legislative acts COM(2021)206 final – 2021/0106(COD), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206> accessed 10 March 2023.

there is a need for more regulation or amendments to the existing regulations, focusing on data protection law and the proposed regulation on AI.

1.3. Research question

The main research question of this thesis is:

> **To what extent is the current EU legal framework for data protection and the proposed artificial intelligence regulation adequate to address discrimination and manipulation of behavior arising from automated decision-making (ADM) used in online platforms?**

The main research question shall be answered through the below-mentioned sub-questions:

i. What is automated decision-making (ADM)?
ii. Why are discrimination and manipulation of behavior problematic in the context of automated decision-making (ADM) used in online platforms?
iii. To what extent are automated decision-making (ADM) and discrimination and manipulation of behavior arising from the use of automated decision-making (ADM) addressed in the current EU data protection regulation and the proposed artificial intelligence regulation?
iv. What could be the possible amendments to the current EU data protection regulation and proposed artificial intelligence regulation to address discrimination and manipulation of behavior arising from automated decision-making (ADM)?

This scope thesis is limited to the following online platforms (i) search engines like Google, (ii) social media platforms including Facebook, Twitter, Instagram, and LinkedIn, (iii) e-commerce marketplaces such as Amazon, (iv) on-demand service providers such as Spotify, Netflix, Youtube, Goodreads. Therefore, while acknowledging the potential harms, the ADM provided as separate software that might be used in various sectors, such as **banking**, **financial services, insurance**, **health care**, **and**, **education** shall be out of the scope of this thesis.

The scope of this thesis is limited to some forms of ADM applications, such as content curation/moderation, targeted and personalized advertisement, and item recommendation and price determination. Other ADM applications are out of the scope of the thesis.

Concerning the legal framework analyzed in this thesis, this analysis is limited to the GDPR and AIA and their relevant provisions for ADM, discrimination and manipulation of behavior. Other potentially relevant frameworks to address discrimination and manipulation of behavior in the context of ADM, such as the anti-discrimination and consumer protection laws, shall be out of the scope of this thesis.

1.4. Methodology

The thesis relies on desk research. The content has been sourced from publicly accessible sources on the Internet; for instance, existing and proposed legislation and relevant Union

body documents are sourced from websites maintained by issuing bodies such as the eur-lex website and edpb.europa.eu websites. Scholarly materials are sourced from books, academic websites such as HeinOnline and ScienceDirect, and articles available at the university library. The search was conducted with a variety of search strings such as "automated decision-making," "online manipulation," "discrimination," "automated decision-making and discrimination," and "automated decision-making and manipulation of behavior." The above-mentioned search strings rendered many articles for each string. The selection of articles was prioritized on classifying discrimination and manipulating behavior and GDPR and AIA analysis of ADM on online platforms.

Content used for this thesis includes the existing data protection rules under the GDPR and the proposed AIA for the analysis, European Union Charter on Fundamental Rights, and the European Convention on Human Rights for defining illegal discrimination definition. The analysis incorporates Union body documents such as Article 29 Working Party (WP29) guidelines, the European Parliamentary Research Service (EPRS) studies, and the European Data Protection Supervisor (EDPS) opinions concerning ADM, manipulation of behavior, and discrimination. Primarily, the analysis used scholarly sources on legal research, general media articles, and studies to analyze the problematic nature of online discrimination and manipulation in online platforms and provide concrete examples. To determine the adequacy of GDPR and AIA, scholarly articles in the context of ADM used in online platforms to address discrimination and manipulation of behavior are used.

1.5. Structure of the Thesis

Chapter 2 of this thesis explains ADM and its underlying technology, namely rule-based and machine learning algorithms. The chapter also focuses on the profiling aspects of ADM. It includes ADM examples used on online platforms that individuals visit. Chapter 2 answers the first sub-question of the thesis.

Chapter 3 of this thesis analyzes discrimination and manipulation of behavior arising from using ADM on online platforms. First, the chapter sets the definition of discrimination and manipulation of behavior. Second, through some examples, it analyzes the problematic nature of discrimination and manipulation of behavior in the context of ADM used on online platforms. Chapter 3 answers the second sub-question of the thesis.

Chapter 4 of this thesis examines how ADM, discrimination, and behavior manipulation are regulated under GDPR and AIA. The chapter analyzes the GDPR and AIA to determine whether ADM, discrimination, and behavior manipulation are addressed adequately. Chapter 4 answers the third sub-question of the thesis.

Chapter 5 of this thesis examines the possible need for further amendments to the GDPR and AIA to address ADM, discrimination, and manipulation of behavior. This chapter provides recommendations for regulations that lack or do not address manipulation of behavior or discrimination arising from using ADM on online platforms. Chapter 5 answers the fourth sub-question of the thesis.

Chapter 6 of this thesis then concludes the analysis to answer the main research question – "to what extent is the current EU legal framework for data protection and the proposed artificial

intelligence regulation adequate to address discrimination and manipulation of behavior arising from automated decision-making used in online platforms?"

2. Defining Automated Decision-Making

This chapter explains automated decision-making (ADM) and the technologies behind such systems: rule-based algorithms and machine learning. It further explains the connection between ADM and profiling and the ADM properties due to the underlying technologies. This chapter, therefore, answers the first sub-question; "What is automated decision-making (ADM)?" Section 2.1. defines ADM, Section 2.2 examines the technology, Section 2.3. analyzes profiling, Section 2.4. discusses ADM properties and Section 2.5 concludes.

2.1. Generic Definitions of ADM

ADM is the process of making an automated decision about individuals without human involvement.[23] Breaking the phrase apart, 'automated' means a system that renders an outcome with automation without any 'major' human involvement; the 'decision-making' part of the phrase provides the decision-making power of the system, including evaluation of metrics or characteristics to provide the outcome. ADM can be based on any data, such as data provided directly from the individual (through surveys on a website), data obtained from the individuals (through data collection), or derived or inferred data that has already been created (through the already existing database).[24]

ADM thus can be seen in many different sectors, such as banking and insurance.[25] An ADM can decide whether an individual receives credit from the bank for their car or house. When an ADM provides a concrete decision about individuals, it is generally more visible to society: "You are eligible for a credit!". ADM on online platforms, in contrast, are not easily acknowledgeable by individuals; often, individuals do not realize that they are subjected to an ADM when seeing an advertisement that impacts their preferences or actions.[26]

2.2. Underlying Technologies of ADM

ADM used in online platforms is based on big data and is generally powered by algorithms and machine learning (ML),[27] which is essential to understand further how ADM works in an

---

[23] 'What Is Automated Individual Decision-Making and Profiling?' (n 6).

[24] Article 29 Data Protection Working Party, 'Guidelines on Automated Decision-Making and Profiling for the Purposes of Regulation 2016/679' (European Commission 2017) WP251rev.01 <https://ec.europa.eu/newsroom/article29/items/612053 > accessed 7 March 2023.

[25] Jeanne G Harris and Thomas H Davenport, 'Automated Decision Making Comes of Age' (MIT Sloan Management Review 2005) <http://pubsonline.informs.org/do/10.1287/8943f842-86f8-4d42-9a64-9a7cd07b31f5/abs/> accessed 20 March 2023.

[26] Moritz Büchi, Eduard Fosch-Villaronga, Christoph Lutz, Aurelia Tamo-Larrieux, Shruthi Velidi, Salome Viljoen 'Chilling Effects of Profiling Activities: Mapping the Issues' [2019] SSRN Electronic Journal <https://www.ssrn.com/abstract=3379275> accessed 20 March 2023.

[27] Ari Ezra Waldman, 'Power, Process, and Automated Decision-Making' (2019) 88 Fordham Law Review <https://ir.lawnet.fordham.edu/flr/vol88/iss2/9/> accessed 7 March 2023.

online platform. To analyze the technology in detail, Section 2.2.1. explains rule-based algorithms, and Section 2.2.2 examines ML algorithms.

2.2.1.   Rule-based Algorithms

Algorithms are defined differently across disciplines – mathematics, computer science, and public discourse.[28] One definition describes an algorithm as *"a sequence of computational steps that transform the input into the output."* [29] Another definition portrays an algorithm as "*a mathematical construct with a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions.*"[30] Despite these definition variations, algorithms can be seen as operational tools that generate an outcome based on an input. There is never one algorithm used for decision-making, but rather a bundle of multiple algorithms with different capabilities.[31] These algorithms often include prioritization, classification, association, and filtering.[32]

Prioritization algorithms use the pre-defined characteristics, metrics, or data fed into the algorithm to rank and sort data in a sorting procedure[33] by prioritizing some metrics and outcomes[34] to create a list. Online platforms use such algorithms to create lists or personalized feeds. For example, the Google search engine ("**Google**") predicts search results by ranking websites,[35] while Netflix or Spotify use it for personalized movie recommendations[36] and music playlists.[37] ***Classification algorithms*** classify the data through a prioritization step by setting a threshold.[38] It enables Google Photos to label "cats"[39] and recognize people[40] or Amazon to

[28] Alan Rubel, Clinton Castro and Adam Pham, 'Algorithms and Autonomy: The Ethics of Automated Decision Systems' (1st edn, Cambridge University Press 2021) <https://philpapers.org/rec/RUBAAA-5> accessed 10 May 2023.

[29] Thomas H Cormen (ed), *Introduction to Algorithms* (3rd ed, MIT Press 2009).

[30] Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951716679679 > accessed 7 March 2023.

[31] Batta Mahesh, 'Machine Learning Algorithms - A Review' (2018) 9 <https://www.ijsr.net/get_count.php?paper_id=ART20203995> accessed 7 March 2023.

[32] Hannah Fry, *Hello World: How to Be Human in the Age of the Machine* (2019).

[33] Nicholas Diakopoulos, 'Algorithmic Accountability Reporting: On the Investigation of Black Boxes' (Columbia Journalism School, TOW Center for Digital Journalism 2013) <https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2> accessed 7 March 2023.

[34] ibid.

[35] 'Ranking Results – How Google Search Works' (*Google Search - Discover How Google Search Works*) <https://www.google.com/search/howsearchworks/how-search-works/ranking-results/> accessed 14 March 2023.

[36] 'Netflix Research - Recommendations' <https://research.netflix.com/research-area/recommendations> accessed 14 March 2023.

[37] 'Amplifying Artist Input in Your Personalized Recommendations — Spotify' <https://newsroom.spotify.com/2020-11-02/amplifying-artist-input-in-your-personalized-recommendations/> accessed 21 June 2023.

[38] Diakopoulos (n 33).

[39] 'How Google Photos Uses Machine Learning to Create Customized Albums' *Christian Science Monitor* <https://www.csmonitor.com/Technology/2016/0324/How-Google-Photos-uses-machine-learning-to-create-customized-albums> accessed 14 March 2023.

[40] 'Recognizing People in Photos Through Private On-Device Machine Learning' (*Apple Machine Learning Research*) <https://machinelearning.apple.com/research/recognizing-people-photos> accessed 14 March 2023.

classify individuals through their browsing or shopping history for advertisements,[41] while YouTube uses it to remove inappropriate videos that breach its Terms and Conditions.[42] **Association algorithms** 'associate' data when a similarity threshold value is met.[43] Amazon uses classification and association algorithms to link individuals with similar purchase histories.[44] **Filtering algorithms** through filtering methods provide accurate, personalized content to individuals[45] enhancing personalized recommender systems, content curation, and targeted advertisements. Facebook often uses filtering algorithms to create a personalized feed with other algorithms.

A comprehensive example would be Amazon's item recommendation system since Amazon uses an *algorithm bundle* for personalized shopping.[46] The system collects personal shopping data, prioritizes items through search (*prioritization algorithm*), classifies individuals based on their interests (*classification algorithm*), associates users with similar purchase histories (*association algorithm*), and filters all items to recommend (*filtering algorithm*).[47] While rule-based algorithms are an option, most companies, including Amazon,[48] prefer ML algorithms for their simplicity, accessibility and effectiveness.[49]

2.2.2. Machine Learning Algorithms

Unlike rule-based algorithms, ML algorithms produce an outcome by creating the rules ("*model*") from the data fed into the system.[50] They apply these rules to new data, learn from their experience, and produce more relevant outcomes.[51] There are three main ML approaches with distinct 'learning' types: supervised, unsupervised, and reinforcement.[52]

Supervised learning is an ML algorithm that learns from pre-determined input classifications.[53] It has *example kit*s, a training dataset, where every input is linked to an output.[54] When the training set has enough input-output pairs, the ML creates a model to link

---

[41] By classification algorithms, can classify individuals through their characteristics or consumption patterns and provide personalized advertisements. *See Kashmir Hill, 'How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did' (Forbes) <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/> accessed 14 March 2023.*

[42] 'Using Technology to More Consistently Apply Age Restrictions' (*blog.youtube*) <https://blog.youtube/news-and-events/using-technology-more-consistently-apply-age-restrictions/> accessed 14 March 2023.

[43] Diakopoulos (n 33).

[44] Fry (n 32).

[45] ibid.

[46] 'The History of Amazon's Recommendation Algorithm' (*Amazon Science*, 2019) <https://www.amazon.science/the-history-of-amazons-recommendation-algorithm> accessed 14 March 2023.

[47] ibid.

[48] 'Recommender System – Amazon Personalize – Amazon Web Services' <https://aws.amazon.com/personalize/> accessed 21 June 2023.

[49] GreekDataGuy, 'Machine Learning Is Getting Easier. Software Engineering Is Still Hard' (*Medium*, 2020) <https://towardsdatascience.com/machine-learning-is-getting-easier-software-engineering-is-still-hard-d4e8320bc046> accessed 21 June 2023.

[50] Mahesh (n 31).

[51] Federico Galli, *Algorithmic Marketing and EU Law on Unfair Commercial Practices*, vol 50 (Springer International Publishing 2022) <https://link.springer.com/10.1007/978-3-031-13603-0> accessed 15 March 2023.

[52] ibid.

[53] Mahesh (n 31).

[54] Galli (n 51).

other inputs with the same metrics embedded in the training set to produce an outcome.[55] The model is tested using a *test dataset* where the ML predicts outcomes by pairing the inputs in the test set, as *Figure 1* shows.[56]
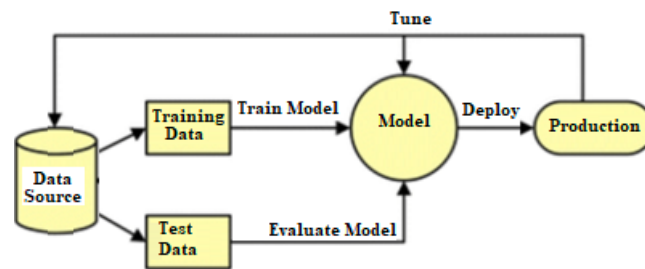
Figure 1: Supervised learning workflow[57]

Supervised learning is a relied-on and popular type of ML algorithm due to its straightforward procedure.[58] It is often used for classification purposes:[59] such as item recommendation in online shopping by analyzing shopping histories and predicting and classifying future shopping preferences.

Unsupervised learning, as the name indicates, does not rely on human supervision or known input-output pairs to train the system.[60] The system operates without a training dataset, sorts the input data, extracts rules, and creates a model to link inputs and their relevant outputs, as illustrated in *Figure 2*.[61]
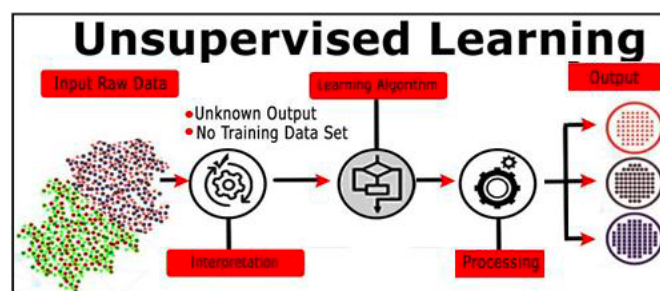
Figure 2: Unsupervised learning workflow[62]

Unsupervised ML is useful for *clustering* techniques[63] to create similar data groups based on system-defined metrics. Such practice benefits online platforms that provide targeted advertisements or use recommender systems that work on similarity thresholds, such as association algorithms.

---

[55] ibid.
[56] ibid.
[57] Mahesh (n 31).
[58] ibid.
[59] Galli (n 51).
[60] ibid.
[61] Mahesh (n 31).
[62] ibid.
[63] Galli (n 51).

Reinforcement learning, on the other hand, does not have a supervisor like unsupervised learning; however, it has a different process. As *Figure 3* shows, the 'agent' learns how to proceed through the feedback from the environment based on its actions and gets 'rewarded' if the problem is solved.[64] In recommender systems, particularly in e-commerce platforms, a reinforcement ML can be used to learn from the consumer interactions to the accuracy of outcomes.[65]
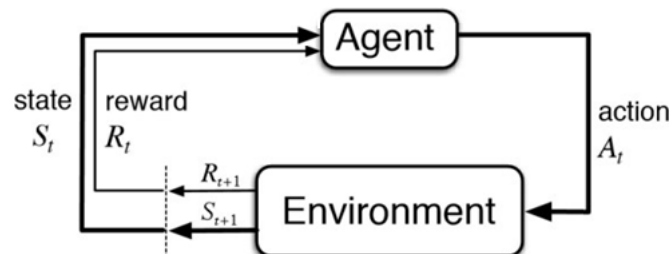
Figure 3: Reinforcement learning workflow[66]

2.3. ADM and Profiling

Profiling is evaluating individuals' characteristics from the gathered data, classifying, or clustering individuals, and determining or predicting their characteristics about individuals.[67] The legal definition of profiling, however, differs between legal scholars. For instance, while some define profiling as "*the systematic and purposeful recording and classification of data related to individuals,*"[68] others provide a more comprehensive explanation, such as "*a process in which characteristics are ascribed to individuals or groups of people, for instance by combining datasets, predicting characteristics or clustering or categorizing people into different groups*."[69] In the digital age, these classifications can now be done through vast databases, leading to more common profiling practices in online platforms.[70]

As its definition suggests, profiling categorizes individuals or groups based on specific characteristics.[71] An ADM with profiling capabilities analyzes, evaluates, predicts, and classifies individuals; in other words, it profiles people on their characteristics more efficiently than human evaluation through automation.[72] While an ADM doesn't need to have profiling

---

[64] Mahesh (n 31).

[65] Galli (n 51).

[66] Mahesh (n 31).

[67] 'What Is Automated Individual Decision-Making and Profiling?' (n 6).

[68] Büchi and others (n 26).

[69] Bart Custers, 'Profiling and Predictions: Challenges in Cybercrime Reasearch Datafication', Anita Lavorgna and Thomas J. Holt (eds), 'Researching Cybercrimes' (1st edn, 2021) < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4016165> accessed 7 March 2023.

[70] Büchi and others (n 26).

[71] Custers (n 69).

[72] Fleur Jongepier and Jan Willem Wieland 'Microtargeting people as a mere means', Fleur Jongepier, Micheal Klenk (eds), *The Philosophy of Online Manipulation* (1st edn, Routledge, 2022) <https://www.taylorfrancis.com/books/9781003205425> accessed 19 March 2023.

activities, ADM in online platforms generally have profiling capabilities to achieve accurate results to serve their commercial benefits.[73]

2.4. Properties of ADM and Their Impact

While ADM provides efficiency for online platforms, it also has different impacts. Section 2.4.1. examines the opacity of the ADM, Section 2.4.2. discusses human involvement in ADM, and finally, Section 2.4.3. examines the privacy invasiveness of ADM.

2.4.1. The Opacity of the System: Invisibility and Black Box Effect

While algorithmic appreciation exists in society,[74] algorithms are often invisible to people.[75] Lack of awareness of being manipulated or discriminated against helps the ADM to achieve its objectives and further aggravates its impacts.[76]

The algorithms are always invisible to individuals in certain contexts,[77] contributing to the system's opacity and understandability. The ADM process is more visible in rule-based algorithms since there are pre-determined rules the algorithm follows. However, in ML, the system extracts the rules and produces an outcome by learning from the environment without pre-determined rules.[78] ML algorithms are often called "*black boxes*" because they do not provide information, rules, metrics, or processes.[79] This lack of transparency affects individuals requesting information on how an outcome is produced concerning them and also the developers of the system, who also might not understand the system they created.[80]

Thus, unlike rule-based algorithms, ML algorithms are harder, or even impossible, to understand due to the lack of pre-determined rules on which the system operates on. Consequently, most black box ML algorithms remain to be unexplainable. While *explainable* algorithms should provide full transparency to the individual, complete transparency may render the algorithm useless in the eyes of its users.[81] Such a lack of explainability of the ADM creates challenges in complying with legal obligations to inform the individuals, hindering transparency and accountability principles that should be complied with.[82]

---

[73] 'What Are Customer Profiling and Its Undeniable Benefits to Business?' (*Quantzig*, 2021) <https://www.quantzig.com/blog/customer-profile/> accessed 21 June 2023.

[74] Jennifer M Logg, Julia A Minson and Don A Moore, 'Algorithm Appreciation: People Prefer Algorithmic to Human Judgment' (2019) 151 Organizational Behavior and Human Decision Processes 90 <https://www.sciencedirect.com/science/article/abs/pii/S0749597818303388> accessed 7 April 2023.

[75] Nick Seaver, 'Knowing Algorithms', Janet Vertesi and David Ribes (eds), *DigitalSTS - A Field Guide for Science & Technology Studies* (Princeton University Press 2019) <https://digitalsts.net/wp-content/uploads/2019/11/26_digitalSTS_Knowing-Algorithms.pdf> accessed 7 March 2023.

[76] Jongepier and Klenk (n 14).

[77] Seaver (n 75).

[78] Fry (n 32).

[79] David Beer, 'Why Humans Will Never Understand AI' <https://www.bbc.com/future/article/20230405-why-ai-is-becoming-impossible-for-humans-to-understand> accessed 21 June 2023.

[80] Seaver (n 75).

[81] ibid.

[82] Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "right to an Explanation" Is Probably Not the Remedy You Are Looking For' (LawArXiv 2017) preprint <https://osf.io/97upg> accessed 21 June 2023.

2.4.2. Human Involvement: Embedded Biases and Errors

The data quality of the input data is a crucial factor in ADM, as "*algorithmic systems are only as good as the data which they are based on.*"[83] In most ADM, human involvement risks carrying and embedding biases into the system at several stages, whether through supervision or active system training.[84] These biases can be explicitly or implicitly introduced into the system leading to the risk of systematic discrimination against certain individuals. Such bias in the training data affects the ADM's fairness and hinders its outcomes' accuracy.[85]

Humans are known to be prone to make errors in their judgments,[86] while computational systems, like ADM, use algorithms designed by humans to make decisions. Although algorithms can typically outperform human judgments,[87] they are not immune to errors in their judgment, either. Such errors are called "*false positive*" or "*false negative*" outcomes and can stem from many reasons, such as incorrect rules[88] or inaccurate or biased training datasets. Such errors might lead to unintentional discrimination or manipulation of behavior.

2.4.3. Privacy Invasive: Manipulation of Behavior

ML algorithms that learn from previous experiences and environments can find unexpected insight (*inferred data*) from the input data.[89] The ability of an ML algorithm to discover information without any specific programming raises concerns about the privacy of the individual. Inferred data can be used to manipulate individuals' behavior through different practices, especially in online platforms through personalized content curation, behavioral targeting practices, and dark patterns.[90]

2.5. Conclusion

This chapter aims to answer the first sub-question; "What is automated decision-making (ADM)?"

ADM are systems that automate decisions about individuals run by either a rule-based or an ML algorithm. Rule-based algorithms operate on pre-determined rules to prioritize, classify, associate, and filter data fed into the system. In contrast, ML algorithms learn from their previous experiences and the environment to create rules and generate more accurate and

---

[83] Waldman (n 27).

[84] Frederik Zuiderveen Borgesius, 'Study on Discrimination, Artificial Intelligence and Algorithmic Decision-Making' (Council of Europe 2018) <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73> accessed 25 April 2023.

[85] Céline Castets-Renard, 'Accountability of Algorithms in the GDPR and beyond: A European Legal Framework on Automated Decision-Making' [2019] SSRN Electronic Journal <https://www.ssrn.com/abstract=3391266> accessed 24 March 2023.

[86] 'Noise: The Other Side Of Human Error' <https://www.forbes.com/sites/hecparis/2021/10/29/noise-the-other-side-of-human-error/> accessed 21 June 2023.

[87] Logg, Minson and Moore (n 74).

[88] Cormen (n 29).

[89] Waldman (n 27).

[90] ibid.

effective outcomes. Online platforms utilize such technologies for content curation, personalized recommendations, and advertisements.

While ADM offers an efficient evaluation of data and decisions, it can have negative consequences on individuals[91] such as errors, technical inexplicability, embedded biases leading to discrimination, and manipulation of behavior. This analysis reveals that while it is possible to intentionally manipulate behavior or discriminate against individuals when explicitly designed, it can also unintentionally create the same results.

The next chapter of this thesis defines discrimination and manipulation of behavior and investigates manipulative and discriminatory practices in ADM used in online platforms.

3. Defining Discrimination and Manipulation of Behavior vis-à-vis Online Platforms

This Chapter defines discrimination manipulation of behavior and discusses the problematic nature of discrimination and manipulation of behavior in the context of automated decision-making used in online platforms. This chapter, therefore, answers the second sub-question; "Why are discrimination and manipulation of behavior problematic in the context of automated decision-making (ADM) used in online platforms?" Section 3.1. defines manipulation, Section 3.2. discusses the problematic nature of manipulation of behavior in the context of online platforms. Section 3.3. defines discrimination, Section 3.4. determines the problematic nature of discrimination in the context of online platforms, and lastly, Section 3.5. concludes.

3.1. Defining Discrimination

Discrimination is the unjust treatment of people or groups based on characteristics such as age, gender, nationality, religion, ethnicity, and sexual orientation, etc.[92] Non-discrimination is a fundamental human right under the Charter of Fundamental Rights of the European Union ("**the Charter**").[93] The Charter prohibits "*discrimination based on sex, race, color, language, religion, political or other opinions, national or social origin, or association with a national minority, property, birth, or another status."[94]* The Charter applies to EU institutions and national authorities when implementing EU law and EU countries when adopting EU directives.[95] In cases where the Charter does not apply, the fundamental right of non-discrimination applies through international conventions ratified by EU countries,[96] such as the

---

[91] Maja Nišević, Alan M. Sears, Eduard Fosch-Villaronga and Bart Custers 'Understanding the Legal Bases for Automated Decision-Making under the GDPR', Eleni Kosta, Ronald Leenes and Irene Kamara (eds), 'Research Handbook on EU Data Protection Law' (1st edn, 2022) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331720> accessed 24 March 2023.

[92] 'Discrimination: What It Is and How to Cope' (*https://www.apa.org*) <https://www.apa.org/topics/racism-bias-discrimination/types-stress> accessed 7 March 2023.

[93] European Union, Charter of Fundamental Rights of the European Union, 26 October 2012, 2012/C 326/02, <https://www.refworld.org/docid/3ae6b3b70.html> accessed 30 March 2023

[94] The Charter, Article 21.

[95] 'When Does the Charter Apply?' <https://commission.europa.eu/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights/when-does-charter-apply_en> accessed 21 June 2023.

[96] ibid.

European Convention on Human Rights ("**ECHR**").[97] The ECHR regulates both direct and indirect discrimination. Direct discrimination involves treating persons in relevant situations based on their characteristics or status,[98] and indirect discrimination occurs when the practice is neutral but becomes discriminatory unintentionally.[99] While the ECHR article does not explicitly mention the application to the private sector, it is interpreted to have a "horizontal effect," applying to purely private matters.[100] All 27 EU Member States are parties to the ECHR, ensuring the application within the EU.[101] Thus, discrimination on online platforms, such as discriminating against older individuals from not seeing job ads on Facebook merely because they are in their later years of life,[102] is almost always considered prohibited discrimination.

The Charter has a limited list of characteristics ("*protected characteristics*"[103]) for prohibited discriminatory practices. Discriminatory practices based on *protected characteristics* are easily determined, as they are prohibited. However, with the advancement of technology and the ability of algorithms to infer data about individuals, online platforms may discriminate against individuals without using any *protected characteristics*.[104] Such discriminatory practices may not be considered illegal under the relevant legislation, as there is no unjust treatment based on *protected characteristics*.

3.2.   Discrimination in Online Platforms

Through the algorithms, ADM can discriminate against people in various practices on online platforms, intentionally or unintentionally. This section analyzes different forms of discrimination practices on online platforms. Section 3.4.1. provides the problematic nature of illegal discrimination, whereas Section 3.4.2. focuses on unfair discrimination on online platforms.

3.2.1.   Illegal Discrimination

Illegal discrimination in ADM used in online platforms is not a new problem brought to light; it dates to the late 2010s when Google's algorithm was accused of being biased toward black women. When searching "black women," Google's prioritization algorithm would prioritize

---

[97] European Convention of Human Rights, Article 14.

[98] 'Guide on Article 14 on the European Convention on Human Rights and on Article 1 of Protocol No.12 to the Convention' (European Court of Human Rights 2022).

[99] European Court of Human Rights, Biao v. Denmark (Grand Chamber), No.38590/10, 24 May 2016, para. 103.

[100] 'Guide on Article 14 on the European Convention on Human Rights and on Article 1 of Protocol No.12 to the Convention' (n 98).

[101] 'European Union Accession to the European Convention on Human Rights - Questions and Answers - Portal - www.coe.int' (*Portal*) <https://www.coe.int/en/web/portal/eu-accession-echr-questions-and-answers> accessed 21 June 2023.

[102] Julia Angwin, Noam Scheiber, Ariana Tobin, 'Dozens of Companies Are Using Facebook to Exclude Older Workers From Job Ads' (*ProPublica*, 2017) <https://www.propublica.org/article/facebook-ads-age-discrimination-targeting> accessed 16 March 2023.

[103] 'Discrimination: Your Rights' (*gov.uk*) <https://www.gov.uk/discrimination-your-rights> accessed 22 March 2023.

[104] Zuiderveen Borgesius (n 84).

explicit adult content on various websites, while no such bias existed for "white women."[105] Although Google's algorithm was said to have been modified since this problem, fixing a specific algorithm did not provide a permanent solution since biased algorithm problems continued to exist within Google. Another instance was Google suggesting advertisements for a *criminal record checking service* when African American-sounding names were searched, while no criminal record-related advertisements were shown for non-African American-sounding names.[106]

Discriminatory practices in ADM are not limited to race or specific online platforms like Google. In an experiment conducted by a non-profit organization[107] in 2020, two fabricated job ads – one for a blue-collar job and one for a white-collar – were displayed on Facebook.[108] While these advertisements were not targeted to a particular gender, the ADM learned from the interactions and started promoting blue-collar jobs for men and white-collar jobs for women, resulting in unintentional discrimination based on gender.[109] When the ADM is not programmed to target the advertisements, it learns gender-stereotypical job pairings from previous interactions due to its ML capabilities and applies them to advertisements that were not targeted.

Meta, which oversees Facebook and Instagram, is known to have "*sensitive, detailed targeting options*" for targeted advertisements on their platforms, including options based on race, ethnicity, religious views, political beliefs, and sexual orientation.[110] These practices create intentional illegal discrimination based on *protected characteristics*, which should be prohibited and illegal. After this covert practice surfaced, Meta-owned online platforms incorporated measures to block advertisers from using such options for their advertisements.[111] However, Facebook, with its massive user database, continues to profile users based on their age and gender, which are also *protected characteristics*, and other data, such as location data for advertisement purposes.[112] Facebook does not necessarily need to intentionally collect protected characteristic data due to the machine learning (ML) algorithms in place for their ADM. The system can infer data or proxies,[113] leading to unintentional discrimination.

---

[105] Noble (n 16).

[106] Damien Gayle, 'Google Accused of Racism after Black Names Are 25% More Likely to Bring up Adverts for Criminal Records Checks' (*Mail Online*, 2013) <https://www.dailymail.co.uk/sciencetech/article-2273741/Google-accused-racism-black-names-25-likely-bring-adverts-criminal-records-checks.html> accessed 24 March 2023.

[107] *'Vision, Mission & Values' (AlgorithmWatch) <https://algorithmwatch.org/en/vision-mission-values/> accessed 22 March 2023.*

[108] Jessica Wulf, 'Guidebook on Automated Decision-Making Systems and Discrimination' (Algorithm Watch 2022) <https://algorithmwatch.org/en/wp-content/uploads/2022/06/AutoCheck-Guidebook_ADM_Discrimination_EN-AlgorithmWatch_June_2022.pdf> accessed 16 March 2023.

[109] ibid.

[110] Emma Roth, 'Facebook and Instagram Will Delete "Sensitive" Ad Targeting Groups Linked to Race, Politics' (*The Verge*, 2021) <https://www.theverge.com/2021/11/9/22773038/meta-detailed-targeting-ads-facebook-instagram-messenger> accessed 23 March 2023.

[111] ibid.

[112] ibid.

[113] Rachel Griffin, 'Tackling Discrimination in Targeted Advertising: US regulators take very small steps in the right direction – but where is the EU?' [2022] Verfassungsblog <https://verfassungsblog.de/targeted-ad/> accessed 23 March 2023.

3.2.2.   Unfair Discrimination

Due to the advancement of technology and the complex capabilities of ADM, new forms of discrimination may arise. These practices may not fall under the illegal discrimination definition but still create "unfair" and problematic practices.[114] For this thesis, unfair discrimination refers to when individuals are discriminated against not based on *protected characteristics* but through other metrics that still result in a discriminatory practice.

As discussed in Chapter 2, ML algorithms can derive information from the training data set or the environment to create rules and unintentionally produce discriminatory outcomes through inferred data and proxies associated with individuals.[115] Some scholars suggest that ADM can only produce intentional discrimination. However, others oppose it and state that it can lead to unintentional discrimination when human bias is embedded into the system.[116] With unintentional discrimination, ADM can pair up internally created proxies with the data within the system to discriminate against individuals based on inferred data or proxies.

Various proxies can be used to infer different characteristics of individuals. Location data is commonly used in advertisements, and it can determine many insights about individuals, such as their neighborhood, which indirectly indicates the individual's socioeconomic status, sexual orientation,[117] or even political beliefs.[118] Location data can also infer individuals' ethnic backgrounds as certain ethnic or religious groups tend to live together in specific neighborhoods, creating very distinct proxies about the race of individuals.[119] With these proxies, online platforms may target X, who resides in a wealthy neighborhood near the port in the Hague, with yacht advertisements, assuming X's wealth and interest in yachts through proxies. However, Y, a yacht enthusiast, may not receive any yacht advertisements because the online platform discriminates against Y based on their location data and classifies him as "poor."[120] Characteristics used in discriminatory practices, such as socioeconomic status or location data used with wealth proxies, do not fall under the *protected characteristics*. Therefore, to claim these advertisements constitute prohibited discrimination, the discrimination grounds should be one of the *protected characteristics*. However, classifying this practice as non-discriminatory may be a "morally" wrong assumption for many individuals.

---

114 Zuiderveen Borgesius (n 84).
115 Anya ER Prince and Daniel Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105 Iowa Law Review < https://ilr.law.uiowa.edu/print/volume-105-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data> accessed 7 March 2023.
116 ibid.
117 Zeynep Tufekci, 'Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency.' [2015] Colo. Tech. LJ 13 <https://heinonline.org/HOL/LandingPage?handle=hein.journals/jtelhtel13&div=18&id=&page=> accessed 20 June 2023.
118 Jon Keegan, 'How Political Campaigns Use Your Phone's Location to Target You – The Markup' (2022) <https://themarkup.org/privacy/2022/11/08/how-political-campaigns-use-your-phones-location-to-target-you> accessed 18 June 2023.
119 Chris Smith, 'Study Reveals the Scary Amount of Personal Data Apps Can Collect without Your Permission' (*BGR*, 2021) <https://bgr.com/tech/location-tracking-personal-data-study/> accessed 18 June 2023.
120 This is a derived example. *See* Prince and Schwarcz (n 115).

Discrimination, nowadays, does not always come with a "one type fits all" approach. One form of unfair discrimination is price discrimination, commonly seen in e-commerce platforms like Amazon. This practice involves a seller changing the product prices for different customers based on available customer data.[121] Amazon uses an ADM to change product prices based on various metrics such as supply and demand or competitor pricing, which are acceptable, but also based on individual shopping history or location data.[122] Although Amazon denied the allegations of engaging in online price discrimination,[123] such practices are examples of unfair discrimination where individuals are discriminated against based on different grounds than *protected characteristics.*

3.3. Defining Manipulation of Behavior

Manipulation is covertly influencing another person's decision-making, while online manipulation specifically refers to using information technology to achieve manipulation.[124] Although manipulation, coercion, and persuasion are often used interchangeably, there are clear distinctions among them. Persuasion is changing individuals' minds through reasons to be reflected and evaluated; coercion restricts the acceptable options individuals might choose.[125] For this thesis, manipulation of behavior refers to imposing an influence over an individual's decision-making in which the individual is unaware of such influence and is steering them to the option the manipulator desires.

The definition of manipulation varies among scholars, but they all share the concept of *hidden influence* over an individual's decision-making autonomy. In traditional and online contexts, manipulation tactics challenge individuals' autonomy and undermine their ability to make independent decisions.[126] In the digital realm, manipulation occurs through ADM that customizes content based on specific metrics, often through profiling, exploiting vulnerabilities, and compromising individuals' autonomy[127] through covert practices. Manipulative online practices can also be facilitated by incorporating dark patterns directing individuals toward specific platform sections.[128] While individuals lack control over their decisions when faced with manipulation, their limited control over the online platform architecture and settings makes them more susceptible to manipulation in an online environment.[129]

---

[121] Kimia Heidary, Bart Custers, Helen Pluut, and Jean-Pierre van der Rest 'A Qualitative Investigation of Company Perspectives on Online Price Discrimination' (2022) 46 Computer Law & Security Review 105734 <https://doi.org/10.1016/j.clsr.2022.105734> accessed 24 March 2023.

[122] Daisuke Wakabayashi, 'Does Anyone Know What Paper Towels Should Cost?' *The New York Times* (2022) <https://www.nytimes.com/2022/02/26/technology/amazon-price-swings-shopping.html> accessed 24 March 2023.

[123] ibid.

[124] Daniel Susser, Beate Roessler and Helen F Nissenbaum, 'Online Manipulation: Hidden Influences in a Digital World' [2018] SSRN Electronic Journal <https://www.ssrn.com/abstract=3306006> accessed 19 March 2023.

[125] ibid.

[126] ibid.

[127] Büchi and others (n 26).

[128] Waldman (n 27).

[129] Jongepier and Klenk (n 14).

3.4.   Manipulation of Behavior in Online Platforms

ADM can manipulate individuals on online platforms, working together with online platforms' architecture.[130] While the online platforms' digital architecture might have a manipulative nature with the embedded dark patterns in the system,[131] this section focuses mostly on the manipulative ADM use in online platforms. Section 3.4.1. explores the different practices of morally acceptable and unacceptable manipulation, and Section 3.4.2 examines intentional and unintentional manipulation seen in online platforms.

3.4.1.   Morally Acceptable and Morally Unacceptable Manipulation

Individuals often welcome certain manipulation of behavior techniques due to the invisibility of the manipulation[132] and the personalization it brings. For instance, individuals may enjoy receiving a recommendation for a movie on Netflix or a second book in a series they have started on their Kindle. The personalization aspect of recommender systems of online platforms is generally embraced by individuals because there is no moral question to it; they are "*morally acceptable*" practices.[133] Amazon predicts the need for a specific item of individuals based on their purchase history, Spotify tailors personalized music lists for individuals, and Twitter and Facebook curate personalized feeds for a personalized user experience. However, manipulative practices can sometimes have an *ulterior motive.* Thus, depending on the context, personalization can also be a "*morally unacceptable*" practice, particularly when combined with *ulterior motives* and widely used online platforms.

The problem of spreading disinformation through personalized content came to light with the Facebook-Cambridge Analytica ("**CA**") scandal, which involved political micro-targeting.[134] CA accessed the data of 50 million Facebook users, created profiles based on individual political views, and sent personalized political messages for Donald Trump's campaign in the 2016 Presidential Election in the United States.[135] The CA identified small groups of people through the characteristics they inferred from Facebook likes, comments, views, and a quiz app that collected data later used to target the individuals with personalized content and advertisement.[136] The content was mostly demonstrable falsehoods,[137] leading to the manipulation of targeted individuals.

---

[130] Thomas Nys, Bart Engelen 'Commercial Online Choice Architecture', Fleur Jongepier, Micheal Klenk (eds), *The Philosophy of Online Manipulation* (1st edn, Routledge, 2022) <https://www.taylorfrancis.com/books/9781003205425> accessed 19 March 2023.
[131] ibid.
[132] Jongepier and Klenk (n 14).
[133] ibid.
[134] ibid.
[135] Carole Cadwalladr and Emma Graham-Harrison, 'Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach' *The Guardian* (2018) <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election> accessed 19 March 2023.
[136] Fry (n 32).
[137] ibid.

Disinformation practices that use ADM are more intricate than anticipated. Dark patterns on online platforms covertly manipulate individuals to click on "secret" advertisements or involuntarily share additional information without knowing it provides more data for manipulation through ADM at a later stage.[138] Targeted individuals in the CA campaign were directed to a quiz app through a dark pattern, which collected personal data to be used in an ADM to produce personalized political posts.[139] CA scandal is a notable example of how *morally unacceptable* it is to manipulate individuals by undermining their decision-making and autonomy. The *ulterior motive* of such practices is not to enhance the personalized user experience but to manipulate political decisions for an upcoming election. The combination of ADM and dissemination of disinformation through personalized feeds[140] creates a problematic form of manipulative ADM practice on online platforms.

The manipulation degree of ADM on online platforms is essential in determining such practices' problematic nature and moral acceptability. Online platforms such as Facebook, Instagram, Twitter, and Google generally rely on personalized, targeted marketing strategies to monetize their platform in exchange for user data.[141] The ADM uses the collected data to generate targeted and personalized advertisements to be displayed on the platform. Additionally, online platforms with personalized recommender systems profile and target individuals with personalized content. While these personalized recommendations may not be as morally wrong as the CA scandal, they still covertly influence individuals and compromise their decision-making abilities and autonomy, adhering to the textbook definition of manipulation. However, one big difference is that disinformation is not a welcomed practice, but receiving a movie recommendation is. The moral acceptability of these practices depends on the context and the *ulterior motive* of the platforms, advertisers, and other actors involved. However, it does not change the manipulative nature of the practice itself.

3.4.2. Intentional and Unintentional Manipulation

Scholars have differing opinions about the intentionality of manipulation. While most scholars argue that manipulation must be intentional to be classified as a manipulation, others suggest that manipulations can be unintentional.[142] The intentionality of the manipulation is closely linked to its *ulterior motive*. In the context of ADM and its algorithms, intentional influences can occur through explicit pre-determined rules, and unintentional consequences may arise because of ADM's actions while not being designed to manipulate individuals intentionally.

---

[138] Jongepier and Klenk (n 14).

[139] Yang Q, 'Algorithms and Resistance: User Behaviors Towards Instagram's Algorithmic Feed' (*Medium*, 2019) <https://medium.com/@yangq.design/algorithms-and-resistance-user-behaviors-towards-instagrams-algorithmic-feed-c090780ee7a0> accessed 20 March 2023.

[140] Geoff Keeling and Christopher Burr 'Digital Manipulation and Mental Integrity', Fleur Jongepier, Micheal Klenk (eds), *The Philosophy of Online Manipulation* (1st edn, Routledge, 2022) <https://www.taylorfrancis.com/books/9781003205425> accessed 19 March 2023.

[141] Ahmed Ginani, 'Why Social Media Platforms Need Your Personal Data' (*Acquaint SoftTech*, 2019) <https://medium.com/acquaint-softtech/why-social-media-platforms-need-your-personal-data-7d23dc115555> accessed 24 March 2023.

[142] Q (n 139).

Intentional manipulation can occur through personalized feeds and targeted advertisements; however, it is not limited to these practices. Experiments on online platforms also constitute an intentional manipulation of behavior, and Facebook is known to experiment with its users regularly. In 2017, investigative journalists discovered that Facebook algorithms could determine the emotions of teenage users and target advertisements toward them.[143] Facebook algorithms can identify when a teenager is feeling "stressed," "stupid," "silly," or a "failure" and allow the advertisers to use this information and target specific advertisements to teenagers.[144] Facebook denied the accusations of targeted advertising and claimed that the experiment was intended to assist advertisers.[145] Regardless of Facebook's denial, this intentional practice exploits vulnerabilities to manipulate individuals, creating a problematic intentional manipulative practice.

Facebook's, or any other online platforms for that matter, ability to predict, determine and exploit an individual's vulnerability through manipulation tactics is no surprise to many. Other online platforms, such as Instagram, Twitter, Pinterest, TikTok, and LinkedIn, use their ADM to create an intentional manipulation medium where the users are generally aware of an algorithmic system's presence but unaware of its manipulation.[146] While these practices theoretically being manipulative, they may not be seen as 'alarmingly' problematic, as mentioned in Section 3.4.1, although the *ulterior motive* and the context matters.

ADM can also unintentionally manipulate, for instance, through ML algorithms behind ADM that manipulate behaviors when not explicitly assigned to do so.[147] The unintentionality of the manipulation through specific practices does not diminish the manipulative effects of ADM but may influence how individuals perceive such practices. Intentional and unintentional practices also risk creating *feedback loops,* such as when similar content based on previous likes and comments is provided to the individual.[148] These expose individuals to the information they already agree with, increasing their biases and creating a *filter bubble* they are unaware exists.[149]

It is straightforward to understand that manipulation of behavior efficiently works together with discrimination; in a way, they complement each other. One of the reasons it works is because manipulation of behavior techniques generally starts by targeting a specific individual or a group through profiling; they exploit the individual's or group's vulnerability and manipulate

---

[143] Susser, Roessler and Nissenbaum (n 124).

[144] Sam Machkovech, 'Report: Facebook Helped Advertisers Target Teens Who Feel "Worthless" [Updated]' (*Ars Technica*, 2017) <https://arstechnica.com/information-technology/2017/05/facebook-helped-advertisers-target-teens-who-feel-worthless/> accessed 20 March 2023.

[145] ibid.

[146] 'Social Media Apps Are "deliberately" Addictive to Users - BBC News' <https://www.bbc.com/news/technology-44640959> accessed 5 June 2023.

[147] Ryan Hogg, 'Elon Musk Called Twitter Newsfeed Algorithms "manipulative," Prompting a Reply from Founder Jack Dorsey, Who Disagreed with the Claim' (*Business Insider*) <https://www.businessinsider.com/musk-and-dorsey-argue-over-manipulative-twitter-algorithms-2022-5> accessed 19 March 2023.

[148] Ricardo Baeza-Yates, 'Personalization, Bias and Privacy.' (2020) <https://dl.acm.org/doi/abs/10.1145/3386392.3399994?casa_token=EwJKdT2wj00AAAAA:fKB9lQNaKGKg7bnVUcGemHAil4sKaGVl7yB45WiJ2o0AALmxMFwbdECQtbXB2lq8gsFga8B48byrQw> accessed 1 July 2023.

[149] Eli Pariser, *The Filter Bubble* (Penguin Books 2011).

them through personalized advertisements or feeds.[150] In this level of intentional manipulation, discrimination plays a significant role in helping online platforms manipulate individual behaviors on a more personal level.

3.5. Conclusion

This chapter aims to answer the second sub-question; "Why are discrimination and manipulation of behavior problematic in the context of automated decision-making (ADM) used in online platforms?"

While discrimination is associated with 'illegal discrimination' based on *protected characteristics*, other forms also exist where they do not fall under these categories. Unfair discrimination occurs when an individual is discriminated against based on not one protected characteristic. Discrimination, regardless of the type, is problematic and morally unacceptable as it involves profiling individuals based on personal characteristics. Manipulation of behavior has a different classification of moral acceptability. Personalized movie or music recommendations, for instance, are *morally acceptable*. In contrast, practices such as disinformation demonstrate *morally unacceptable* practices. While online platforms generally intentionally discriminate and manipulate individuals, there can also be instances without such intent due to the abilities of ADM.

This analysis shows that while discrimination is always problematic, the problematic nature of behavior manipulation relies on the *ulterior motive* and the context. It may be challenging to regulate *all* discriminatory practices regardless of the protected characteristics and *some* problematic manipulative practices depending on their corresponding context.

The next chapter shall focus on how the EU legal framework addresses these issues.

4. EU Data Protection and Artificial Intelligence Legislations

*This chapter focuses on how the General Data Protection Regulation ("**GDPR**") and the proposed Artificial Intelligence Act ("**AIA**") address the discrimination and manipulation of behavior of automated decision-making (ADM) used on online platforms. This chapter, therefore, answers the third sub-question; "To what extent are automated decision-making (ADM) and discrimination and manipulation of behavior arising from the use of automated decision-making addressed in the current EU data protection regulation and the proposed artificial intelligence regulation?". To determine the specific regulation methods, Section 4.1 introduces the EU's general approach to regulating data protection and artificial intelligence, Section 4.2. and 4.3. analyze the GDPR and AIA in detail, respectively, and Section 4.4. concludes.*

---

[150] Jongepier and Klenk (n 14).

4.1. The EU's "Similar-but-Different" Approach to Regulate Data Protection and Artificial Intelligence

The GDPR is among the most substantial legislative instruments for personal-data-driven digital technologies.[151] Like its predecessor, Directive 95/46/EC,[152] it has incorporated a risk-based approach, in line with the opinions of Article 29 Working Party ("**WP29**").[153] The proposed AIA similarly adopts a risk-based approach, although the two regulations differ significantly in their approaches to risk mitigation.[154] The GDPR focuses on ensuring compliance with data protection requirements,[155] while the AIA classifies the risk levels of AI systems.[156] While both are risk-based approaches, they have different regulatory models and serve different aims.[157] These legal instruments thus have different impacts on the individuals affected by manipulation and discrimination through the ADM in online platforms.

4.2. The EU's Data Protection Approach

The GDPR aims to strengthen individuals' fundamental rights through clear rules for the EU digital single market while imposing obligations for data controllers.[158] The GDPR's material scope is broad, as it applies to all data processing wholly or partly by automated means.[159] It provides general principles and rules for data processing rather than focusing on the specific impact of data processing practices.

While all data processing practices are subjected to these general provisions, ADM, on the other hand, is subjected to specific requirements. ADM is regulated under Article 22(1) GDPR, which states, "*the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.*"

While not an essential requirement of ADM, profiling is defined as "*any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular, to analyze or predict aspects concerning*

---

[151] European Council, 'The General Data Protection Regulation' (2022) <https://www.consilium.europa.eu/en/policies/data-protection/data-protection-regulation/> accessed 5 June 2023.

[152] Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data 1995 [95/46/EC].

[153] 'Statement on the Role of a Risk-Based Approach in Data Protection Legal Frameworks' (Article 29 Data Protection Working Party 2014) WP 218 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf> accessed 23 April 2023.

[154] Raphaël Gellert, 'The Role of the Risk-Based Approach in the General Data Protection Regulation and in the European Commission's Proposed Artificial Intelligence Act: Business as Usual?' (2021) 3 <https://jelt.padovauniversitypress.it/2021/2/2> accessed 28 May 2023.

[155] 'Statement on the Role of a Risk-Based Approach in Data Protection Legal Frameworks' (n 153).

[156] 'A European Approach to Artificial Intelligence | Shaping Europe's Digital Future' (n 20).

[157] Gellert (n 154).

[158] 'Data Protection in the EU' (2021) <https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en> accessed 28 May 2023.

[159] GDPR, Article 2(1).

that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements."[160]

The GDPR's approach to regulating ADM is worth analyzing since it provides "**a right not to be subjected to ADM**." While there are debates on whether Article 22 GDPR only applies when invoked as a data subject right ("**DSRs**"),[161] Article 29 Working Party ("**WP29**") sets a straightforward interpretation: Article 22(1) GDPR states a <u>**general prohibition**</u> for decisions solely based on ADM.[162] This prohibition applies when the decisions create a 'legal' or 'similarly significant' effect, but these effects are not defined in the GDPR.[163] WP29 describes legal effects as decisions that affect someone's legal rights or status.[164] Similarly significant effects refer to a similar effect to a legal effect, such as decisions that might affect an individual's access to health services, education, or financial situation.[165] To further explain this concept, WP29 gives an example of targeted advertisements that do not have significant effects; however, depending on the context, it may also have a significant effect, i.e., when an advertisement targets individuals based on their vulnerabilities.[166]

While a prohibition is set forward, Article 22(2) GDPR introduces some exemptions to this rule. These exemptions include the decision is necessary for entering into or performing a contract between the data subject and a data controller; authorized by Union or Member State law to which the controller is subject and with suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or based on the explicit consent of the data subjects.[167]

While the GDPR refrains from defining <u>explicit</u> consent, it raises concerns over the consent-based exemption, which may reduce individuals' de facto level of protection, contrary to the aim of the provision.[168] Another exemption that may create issues in the context of ADM is the Union or Member State law authorizing ADM use. The article does not provide further information or limit the application of such exemption to specific practices; however, some examples are provided in the corresponding recital, such as tax fraud monitoring.[169] The article's lack of clarity might lead to fragmentation in the harmonized rules for ADM, inconsistent practices, and practical discrepancies within the EU digital single market if used by online platforms.[170] Despite the decrease in protection levels and interpretational matters,

---

[160] GDPR, Article 4(4).

[161] Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond' (2019) 27 International Journal of Law and Information Technology 91 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3124901> accessed 28 May 2023; Article 29 Data Protection Working Party (n 24).

[162] Article 29 Data Protection Working Party (n 24).

[163] GDPR, Recital 71.

[164] ibid.

[165] Article 29 Data Protection Working Party (n 24).

[166] ibid.

[167] GDPR, Article 22(2).

[168] Isak Mendoza and Lee A Bygrave, 'The Right Not to Be Subject to Automated Decisions Based on Profiling' in Tatiana-Eleni Synodinou and others (eds), *EU Internet Law* (Springer International Publishing 2017) <http://link.springer.com/10.1007/978-3-319-64955-9_4> accessed 18 May 2023.

[169] GDPR, Recital 71.

[170] Brkan (n 161).

these exemptions open the '*legal'* pathway for ADM use in their recommender systems in online platforms through the explicit consent exemption.

Article 22(3) GDPR requires data controllers to implement measures to safeguard DSRs, freedoms, and legitimate interests when an ADM is being used. The obligation includes a measure to ensure the right to obtain human intervention, to express their point, and to contest the decision when ADM is used.[171] The article does not mention specific safeguards; however, the recital states several measures. These measures include using mathematical or statistical procedures for profiling, implementing technical and organizational measures to correct inaccuracies and minimize the risk of inaccurate decisions, and securing personal data to prevent discriminatory effects on individuals by considering potential risks.[172] Although not a critical tool to ensure consistent compliance with the GDPR,[173] WP29 has provided non-exhaustive good practice recommendations that, most importantly, include: regular quality checks of the ADM to ensure fair treatment and prevent discrimination; algorithmic auditing to ensure the algorithms perform as intended and to prevent discriminatory, erroneous, or unjustified results; third-party auditing for ADM based on profiling has a high impact, ethical review boards to assess potential harms of profiling-based ADM.[174]

While the GDPR does not explicitly address discrimination and manipulation of behavior arising from ADM, some tools can help individuals with manipulative or discriminatory ADM practices: Article 22, Recital 71 works together with data subject rights ("**DSRs**") under Articles 13-15.[175] Articles 13 and 14 GDPR (*right to information*) and Article 15 GDPR (*right to access*) regulate the information provided to the individual. Both DSRs explicitly mention Article 22(1) GDPR and require that if an ADM is present, the individual is to be provided with "*meaningful information about the logic involved as well as the significance and the consequences of such processing.*"[176]

While both DSRs look similar concerning their impact on the explainability of ADM, they serve different purposes and offer different protections.[177] While Articles 13 and 14 provide *ex-ante* preliminary information to fulfill a data controller obligation, Article 15 offers *ex-post* information for the individual when the right is invoked.[178] However, the GDPR's approach to the explainability of ADM has been extensively criticized. Some argue that applying these rights

---

[171] GDPR, Article 22(3).
[172] GDPR Recital 71.
[173] Margot E Kaminski and Gianclaudio Malgieri, 'Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations' (2021) 11 International Data Privacy Law <https://academic.oup.com/idpl/article/11/2/125/6024963> accessed 28 May 2023.
[174] Article 29 Data Protection Working Party (n 24).
[175] Jersain Zadamig Llamas Covarrubias, 'Algorithmic Transparency as a Foundation of Accountability' [2022] SSRN Electronic Journal <https://www.ssrn.com/abstract=4160223> accessed 3 June 2023.
[176] GDPR, Article 13(2)(f), 14(2)(g) and 15(1)(h).
[177] Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 243 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3088976> accessed 28 May 2023.
[178] ibid.

may be limited in cases where minimum human intervention is present.[179] Some underline these articles' vague and broad nature, which may lead to interpretational matters concerning their practical application and enforcement by data controllers and national authorities.[180]

4.3. The EU's Artificial Intelligence Approach

The EC presented the AIA proposal ("**Commission Draft**") on 21 April 2021 in line with its European AI Strategy.[181] The AIA aims to address the risks generated by AI systems through a risk-based approach, classifying different AI systems that impact individuals' safety, health, and fundamental rights.[182] The European Parliament ("**EP**") has adopted its negotiating position on 14 June 2023, sending it to the Council of the EU for trilogues ("**Parliament Draft**").[183] The EP aims to ensure that AI systems are safer, more transparent, traceable, and non-discriminatory than the Commission Draft.[184]

Unlike the GDPR, AIA focuses on AI systems rather than regulating ADM. The AIA applies to AI systems, defined as machine-based systems that operate with varying levels of autonomy and generate outputs such as predictions, recommendations, and decisions that influence physical or virtual environments.[185] The AIA has a four-tiered risk approach: unacceptable risk, high-risk, limited risk, and low or minimal risk.[186] AI systems with unacceptable risk categories are fully prohibited, while the other risk categories have certain obligations to fulfill, such as high-risk AI systems. To analyze the regulation in detail, Section 4.3.1. analyzes prohibited AI systems, Section 4.3.2. analyzes high-risk AI systems and Section 4.3.2. examines general obligations for all AI systems.

4.3.1. Prohibited AI Systems

Article 5 AIA regulates the 'unacceptable risk' category. It provides the prohibited practices of *manipulation:*

    a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness ***or purposefully manipulative or deceptive techniques, to or the effect of*** materially distorting a person's or a group of persons' behavior by appreciably ***impairing the person's***

---

[179] Sandra Wachter, Brent Daniel Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' [2017] International Data Privacy Law <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469> accessed 28 May 2023.

[180] Malgieri and Comandé (n 177).

[181] 'MEPs Ready to Negotiate First-Ever Rules for Safe and Transparent AI | News | European Parliament' (2023) <https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai> accessed 16 June 2023.

[182] 'A European Approach to Artificial Intelligence | Shaping Europe's Digital Future' (n 20).

[183] 'MEPs Ready to Negotiate First-Ever Rules for Safe and Transparent AI | News | European Parliament' (n 181).

[184] 'AI Act: A Step Closer to the First Rules on Artificial Intelligence | News | European Parliament' (2023) <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> accessed 16 May 2023.

[185] Parliament Draft of AIA, Article 3(1).

[186] 'Regulatory Framework Proposal on Artificial Intelligence | Shaping Europe's Digital Future' (2023) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 4 June 2023.

ability to make an informed decision, thereby causing the person to take a decision that that person would not have otherwise taken in a manner that causes or is likely to cause that person, another person or group of persons **significant harm**; The prohibition of an AI system that deploys subliminal techniques referred to in the first subparagraph shall not apply to AI systems intended to be used for approved therapeutical purposes based on specific, informed consent of the individuals that are exposed to them or, where applicable, of their legal guardian;

b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a person or a specific group of persons, *including characteristics of such person's or a such group's known or predicted personality traits or social or economic situation* age, physical or mental ability with the objective or to the effect of materially distorting the behavior of that person or a person pertaining to that group in a manner that causes or is likely to cause that person or another person **significant harm**;

While both points a and b are interpreted as "manipulative AI practices"[187] to fit better with the structure of this thesis, Article 5, paragraph 1 point a ("**manipulative AI**") and Article 5 AIA, paragraph 1, point b ("**discriminatory AI**") shall be analyzed differently.

Commission Draft does not explicitly mention "manipulation" in Article 5 AIA for manipulative AI. However, Recital 15 mentions that AI systems that are "*misused or provide powerful tools for manipulative, exploitative and social control practices*" should be prohibited because of the harm imposed on society.[188] In contrast, the Parliament Draft explicitly incorporates "*purposefully manipulative or deceptive techniques*" in Article 5 AIA. The EP also explicitly includes the elements of manipulation in the article to further emphasize the intent of such systems that alter an individual's autonomy to make informed decisions.[189]

The prohibition in Article 5 AIA applies only in instances where manipulation is *intentional*. From an online platform perspective, unintentional outcomes are also possible with recommender systems, although the prohibited practices do not cover such outcomes. Moreover, the current prohibitions cover practices that online platforms intentionally use, including content moderation or personalized recommendations which are manipulative. Thus, the effectiveness of Article 5 AIA in the context of online platforms may be limited.[190]

The Commission Draft proposed a limited approach to vulnerabilities for discriminatory AI, focusing on "*vulnerabilities of a specific group of persons,*"[191] leaving a worrying gap for other

---

[187] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 4/2021 Computer Law Review International <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3896852> accessed 28 May 2023.
[188] Commission Draft of AIA, Recital 15.
[189] Parlimanet Draft of AIA, Article 5.
[190] Veale and Zuiderveen Borgesius (n 187).
[191] Commission Draft of AIA, Article 5(1)(b).

protected characteristics, including the individual-based ones.[192] In contrast, the Parliament Draft expanded the article to include a "*person or a specific group of persons,*" incorporating individual characteristics while explicitly addressing limited vulnerabilities.[193]

Recital 16a in the Parliament Draft recognizes the importance of prohibiting discriminatory AI systems that categorize individuals based on known, inferred, or *protected characteristics* such as gender, gender identity, race, ethnic origin, sexual orientation, and religion, citing Article 21 of the Charter of Fundamental Rights of the European Union.[194] However, important *protected characteristics* such as race, color, ethnic origin, or sexual orientation are still not explicitly mentioned in the article, continuing the concern voiced for the Commission Draft. While the addition of Recital 16a is a good step in acknowledging different vulnerabilities, the imbalance between the recital and the article needs further discussion and evaluation. Although there is unclearness concerning illegal discrimination, the Parliament Draft incorporates discrimination beyond *protected characteristics*, such as social or economic situation, in Article 5 AIA.[195] This broadens the interpretation of Article 5(1)(b) AIA to include all forms of discrimination, including *unfair discrimination*.

Apart from specific discussions, Article 5 in the Commission Draft was criticized for having vague language and an outdated scope compared to existing legislation in the other fields.[196] The Parliament Draft provided a more in-depth prohibition article; however, the interpretational problems remain, especially with the "significant harm" requirement for prohibited AI systems.

4.3.2.   High-Risk AI Systems

The Commission Draft of the AIA did not address online platforms or recommender systems. However, The Parliament Draft acknowledges the emergence of social media and its strong influence over safety, shaping public opinions, elections, and societal concerns, and the need for AI systems used in the recommender systems to be regulated under the AIA.[197] The Parliament Draft, therefore, has amended Annex III, paragraph 1, point 8 to include the recommender systems of online platforms as a high-risk AI system:

> a b) *AI systems intended to be used by social media platforms designated as very large online platforms within the meaning of Article 33 of Regulation EU 2022/2065 in their recommender systems to recommend to the recipient of the service user-generated content available on the platform.*

---

[192] 'EU's AI Law Needs Major Changes to Prevent Discrimination and Mass Surveillance' (*European Digital Rights (EDRi)*) <https://edri.org/our-work/eus-ai-law-needs-major-changes-to-prevent-discrimination-and-mass-surveillance/> accessed 4 June 2023.

[193] Parliament Draft of AIA, Article 5(1)(b).

[194] Parliament Draft of AIA, Recital 16a.

[195] Parliament Draft, Article 5(1)(b).

[196] Veale and Zuiderveen Borgesius (n 187).

[197] Parliament Draft AIA, Recital 40a.

This addition ensures that recommender systems of online platforms that are classified as "very-large-online-platforms" ("**VLOPs**") under the Digital Services Act ("**DSA**")[198] comply with the requirements for high-risk AI systems. These requirements include:

- *the risk management system* that shall be used throughout the high-risk AI system's lifecycle to identify and analyze known and foreseeable risks, estimate and evaluate risks when the system is used with its intended purpose, adopt suitable risk management measures,
- *data and data governance* measures that require the datasets used for the training, validation, and testing of the AI system to have specific requirements such as being relevant and free of errors,
- *technical documentation* measures that demonstrate compliance with the AIA and provide the necessary information,
- *record-keeping (logging)* measures to log all events and have traceability of the AI system's functions throughout its lifecycle,
- *transparency* measures to ensure the transparency of the operation to enable the users to interpret the system's outputs and obtain transparent information, including the intended purposes of the system,
- *human oversight* measures that ensure natural persons effectively oversee the system,
- *accuracy, robustness, and cybersecurity* measures that ensure the consistent performance of the system with the appropriate level of accuracy, robustness, and cybersecurity.[199]

Including recommender systems within the scope of the AIA ensures consistent checks on the intended use of the system, implementation of risk management measures, and the use of inclusive and error-free datasets and transparent information that can be given to the users concerning the AI system, including its intended use.[200]

The Parliament Draft introduces another high-risk AI system that mitigates a problematic manipulatory practice. This addition entails AI systems directly interacting with natural persons that intend to influence the outcome of an election or referendum or individuals' voting behavior while exempting structural AI systems used for organizational purposes in political campaigns.[201] This addition acknowledges one of the *morally unacceptable* practices seen in online platforms by recognizing its problematic nature.

4.3.3. General Obligations for all AI Systems

The Parliament Draft follows a similar approach to the GDPR by incorporating general principles applicable to all AI systems.[202] Although these principles are prescribed as

---

[198] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) OJ L 277/1 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065&qid=1688059965544> accessed 10 June 2023, Article 33.
[199] Parliament Draft AIA, Articles 8-15.
[200] ibid.
[201] Parliament Draft AIA, Annex III, sub-paragraph 8(ab).
[202] Parliament Draft AIA, Article 4a.

requirements for high-risk AI systems under Articles 8-15 AIA,[203] it incorporates additional explanations for each principle, such as privacy, transparency, non-discrimination, and fairness. These principles ensure that AI systems comply with data protection rules,[204] allow traceability and explainability,[205] and prevent discriminatory impacts and unfair biases prohibited by Union or national laws.[206]

The AIA imposes additional transparency obligations for AI systems that interact with individuals, requiring individuals to be informed that they are interacting with an AI system unless it is obvious from the context of use.[207] The Parliament Draft expands on this requirement by specifying information that should be provided to individuals, including the system's functions, the presence of human oversight, and individuals' existing rights of the individual under EU or national law, including the "right to seek an explanation,"[208] enhancing the individual protection levels.

4.4. Conclusion

This chapter focused on answering the third sub-question; "To what extent are automated decision-making (ADM) and discrimination and manipulation of behavior arising from the use of automated decision-making addressed in the current EU data protection regulation and the proposed artificial intelligence regulation?".

The GDPR and the AIA have risk-based approaches but differ in regulating specific problems. The GDPR prohibits ADM explicitly, including profiling, under Article 22(1) GDPR, with exemptions provided under Article 22(2) GDPR, such as the individual's explicit consent and implementing measures to safeguard the individuals' fundamental rights under Article 22(3) GDPR. However, the GDPR does not explicitly address manipulative or discriminatory ADM. Nevertheless, some tools provide power to individuals through the right to information and access under Articles 13-15 GDPR to address such practices.

Conversely, the AIA focuses on AI systems that can generate recommendations, predictions, and decisions.[209] The AIA explicitly prohibits manipulative and discriminatory AI systems if significant harm occurs to the user. Unlike the GDPR, the AIA also recognizes recommender systems in online platforms as high-risk AI systems and imposes obligations on the online platforms classified as VLOPs under the DSA.

The next chapter evaluates possible amendments for both legislations to increase the level of protection provided to individuals.

---

[203] Parliament Draft AIA, Article 4a(2).
[204] Parliament Draft AIA, Article 4a(1)(c).
[205] Parliament Draft AIA, Article 4a(1)(d).
[206] Parliament Draft AIA, Article 4a(1)(e).
[207] Parliament Draft AIA, Article 52(1).
[208] Parliament Draft AIA, Article 52(1), subparagraph 2.
[209] AIA, Article 3(1).

5. Possible Improvements to the EU's Legislative Approaches

*This chapter investigates possible amendments to the General Data Protection Regulation ("**GDPR**") and the Artificial Intelligence Act ("**AIA**") to better address the manipulation and discrimination of automated decision-making (ADM) in online platforms. This chapter, therefore, answers the fourth sub-question; "What could be the possible amendments to the current EU data protection regulation and proposed artificial intelligence regulation to address discrimination and manipulation of behavior arising from the use of automated decision-making (ADM)?". To determine the possible improvements, Section 5.1. focuses on possible amendments to the GDPR, Section 5.2. determines possible amendments to AIA, Section 5.3. provides other recommendations, Section 5.4. concludes.*

5.1. Analysis and Four Amendments to the GDPR

The GDPR does not have specific rules to regulate the impact of ADM, such as discrimination and manipulation of behavior arising from using ADM in online platforms, resulting in legal uncertainties when mitigating specific outcomes.[210] However, in a highly data-driven ecosystem, the GDPR can have a positive impact by being interpreted and subjected to amendments to enable useful applications of AI, to ensure legal certainty and protect data subjects.[211] This section analyzes and provides four amendments to the GDPR. Section 5.1.1. provides two recommendations for Article 22 GDPR to ensure legal certainty, Section 5.1.2. provides two recommendations for the 'uncertainty' of Articles 13, 14, and 15 GDPR to empower data subject rights ("**DSRs**").

5.1.1. Clarity to Article 22 GDPR

As set forward in Chapter 4 of this thesis, the primary approach of the GDPR is the prohibition of using ADM regardless of their contextual use.[212] However, the wording of Article 22 GDPR is considered ambiguous and lacks clarity on certain aspects despite having a general focus on ADM.[213] [214]

Firstly, one ambiguity of Article 22 GDPR is the lack of clarity over concepts of "prohibition," "legal or similarly significant effect," and "explicit consent." Although intended as a prohibition, the article is written as a '**right not to be subjected to ADM**.'[215] The complex wording creates

---

[210] 'Study on The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (European Parliamentary Research Service 2020) <https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530> accessed 1 June 2023.

[211] 'Study on The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (n 210).

[212] GDPR, Article 22(1).

[213] Article 29 Data Protection Working Party (n 24).

[214] 'Study on The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (n 210).

[215] GDPR, Article 22(1)

applicability problems in the practical world. Therefore, Article 22 GDPR should be reworded to establish a clear legal prohibition on using ADM. [216]

Article 22(1) also lacks clarity on the term "**legal or significant effect**." The article does not provide further explanation about legal and similar significant effects. In contrast, the corresponding recital offers limited examples without any enlightening explanation on understanding such effects. Article 29 Working Party ("**WP29**") also interprets that only 'serious impactful effects' fall under Article 22 GDPR.[217] However, due to the general approach of the GDPR and the lack of examples, it is practically challenging to comprehend the 'serious impactful effects' caused by ADM. While some instances may have minimal effects, there are numerous instances where an ADM can manipulate or discriminate against individuals. Therefore, it is necessary to incorporate a legally binding definition of "legal and similar significant effect."[218] This amendment could align with the WP29 opinion, which adopts a contextual interpretation for targeted advertisement.[219]

Another uncertainty lies with the '**explicit consent**' exemption for using ADM under Article 22(2) GDPR. The GDPR, albeit requiring explicit consent, does not define '*explicit*' consent or determine a distinction from 'consent.' A clear definition of explicit consent should be added in Article 4 GDPR to address this. This amendment should establish that explicit consent requires an 'expressly confirm consent'[220] rather than a 'clear affirmative action' in line with WP29 opinions.[221]

Secondly, another recommendation is to enhance the safeguards under Article 22(3) GDPR since the GDPR focuses on individual entitlements but not on the impacts of ADM. Thus, it is crucial to incorporate legally binding safeguards.[222] Article 22(3) GDPR does not elaborate on the specific safeguards to be implemented, and neither does the general obligation to ensure appropriate technical and organizational measures in Article 24 GDPR. Considering AI implications under the ADM, such safeguards have increased importance over discrimination, manipulation of behavior, explainability, accountability, and transparency.[223]

The absence of explicit recognition of the DSRs in Article 22 GDPR also raises concerns, and it is necessary to introduce a clear obligation, a right to explanation under Article 22(3), as a safeguard, among other measures. Such an amendment can ensure ADM processes' explainability, accountability, and transparency.[224]

---

[216] 'Radical Rewriting of Article 22 GDPR on Machine Decisions in the AI Era' (*European Law Blog*, 2021) <https://europeanlawblog.eu/2021/10/13/radical-rewriting-of-article-22-gdpr-on-machine-decisions-in-the-ai-era/> accessed 18 June 2023.

[217] Article 29 Data Protection Working Party (n 24).

[218] Wachter, Mittelstadt and Floridi (n 179).

[219] Article 29 Data Protection Working Party (n 24).

[220] Article 29 Working Party, 'Guidelines on Consent under Regulation 2016/679' (2017) WP259 <https://ec.europa.eu/newsroom/article29/items/623051/en> accessed 1 June 2023.

[221] GDPR, Article 4(11).

[222] 'Study on The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (n 210).

[223] 'Radical Rewriting of Article 22 GDPR on Machine Decisions in the AI Era' (n 216).

[224] Wachter, Mittelstadt and Floridi (n 179).

It is necessary to amend Article 22(3) GDPR to incorporate the following <u>minimum</u> requirements: [225] [226]

    (i)   regular control over the datasets and proxies,
    (ii)   the identification of biases and unfairness that might lead to discrimination,
    (iii)   periodic algorithmic audits,
    (iv)   implementation of measures for data minimization,
    (v)   ethical review boards to assess potential harms to individuals,
    (vi)   DSRs as outlined under Articles 13,14 and 15 GDPR for increased protection.

Thus, there can be two amendments to Article 22 GDPR:

> - The clarity to the "prohibition" element and "legal or similar significant effect" in Article 22(1) and "explicit consent".
> - Improvement of individual protection levels by incorporating minimum safeguards including the addition of DSRs.

5.1.2.   Clarity to Articles 13, 14, and 15 GDPR

The DSRs under Articles 13, 14, and 15 are the only legally binding obligations for the online platforms to provide information about the ADM that helps protect individuals against discrimination and manipulation of behavior, although some ambiguity exists.

Firstly, "meaningful information" and "envisaged consequences" terminologies are vague. The GDPR acknowledges the importance of transparency, especially when the system's complexity makes it difficult for the individual to understand the process.[227] However, due to the complexity of ADM technology, providing "meaningful" information about the logic or the "envisaged consequences" of the processing may not always be feasible. To address this, there should be clear definitions of "meaningful logic" and "envisaged consequences" in relevant articles or Article 4 GDPR. The "meaningful logic" definition should include a comprehensive explanation of ADM processes and the metrics used, in line with the WP29 opinion.[228] Similarly, the "envisaged consequences" definition should include the possible impacts on the individual, including associated risks.

Secondly, the wording of Articles 13-14 (*right to information*) and Article 15 (*right to access*) of the GDPR is identical, requiring "meaningful information about the logic involved" and "envisaged consequences of the processing" to be provided to the individual.[229] There should be a clear distinction between the explanations provided to the individual in different situations. It is important to determine whether Article 15 GDPR, like Articles 13-14 GDPR, provides general information about ADM or it provides user-centric and specific information. While the WP29 suggests that the controller should provide general information under Article 15 GDPR,[230] there should be a clear difference between *ex-ante* and *ex-post* information.

---

[225] 'Study on The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (n 210).
[226] Article 29 Data Protection Working Party (n 24).
[227] GDPR, Recital 58.
[228] Article 29 Data Protection Working Party (n 24).
[229] GDPR, Article 13(2)(f), 14(2)(g) and 15(1)(h).
[230] Article 29 Data Protection Working Party (n 24).

When individuals invoke their right of access to obtain information about ADM processes that manipulate or discriminate against them, relying on the interpretation of WP29 would limit the information provided to individuals by not providing situation-based information, which goes against the nature of the right. Individuals discriminated against or manipulated in an online platform should be able to obtain specific information about the process that caused harm to them. Therefore, rewriting Article 15 GDPR to include elements for situation-based information rather than general information is strictly necessary.

Thus, two amendments could be helpful for the practical application of the DSRs:

> - Clarity for "meaningful information about the logic involved" and "envisaged consequences of the processing" in Articles 13,14 and 15 GDPR.
> - Clear distinction between *ex-ante* and *ex-post* information, by including a user-centric and situation-based transparency requirement under Article 15 GDPR.

5.2. Analysis and Five Amendments to the AIA

While the AIA is still in negotiation, five improvements can be made to have a clear and effective AIA regulating specific AI systems problems. Sections 5.1.1. discusses clarity for the scope of the AIA, Section 5.1.2. examines the lack of definitions for recommender systems, Section 5.1.3. analyzes the classification of online platforms in high-risk AI systems. Section 5.1.4. examines significant harm requirement, and Section 5.1.5. discusses the feasibility of some high-risk AI system requirements.

5.2.1. Clarity for AI System Definition

The European Commission's Draft ("**Commission Draft**") defined AI systems in Annex I as software that incorporates technologies such as machine learning approaches, logic- and knowledge-based approaches, and statistical approaches which explicitly mentioned rule-based algorithms, and supervised, unsupervised and reinforcement learning.[231] However, these technological techniques were not defined within the body of the AIA, which goes against Recital 6 AIA, which specifically mentions that AI systems should be clearly defined to ensure legal certainty and provide flexibility to accommodate future AI field evolution.

In their General Approach, the Council of the European Union ("**Council**") recommended the deletion of Annex I, incorporating the approaches in Article 3 AIA, and adding new recitals to clarify what machine-learning and logic- and knowledge-based approaches meant for the interpretation of the AIA.[232] The European Parliament's Draft ("**Parliament Draft**") removed Annex I and revised the definition of the AI system. While this change aligns with Council's approach, it provides a more general scope. With the Commission Draft, the specific mention

---

[231] Commission Proposal AIA, Annex I.
[232] Council of EU, 'General Approach on Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' (Council of the European Union 2022) 2021/0106(COD) <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf> accessed 1 June 2023.

of logic- and knowledge-based approaches had potential interpretational issues; the Parliament Draft definition covers all types of technologies that make predictions, recommendations, or decisions.[233] The impact of this amendment on legal certainty can be debated, but the current definition ensures a technology-neutral nature, decreases interpretational issues, and minimizes potential loopholes.

Although it is important to ensure technology-neutral wording, it is also important to understand the impacts of these technologies to comply with the requirements better. A combination of the Commission Draft and the Parliament Draft approaches would be ideal to achieve a more comprehensive and beneficial scope for AI systems. The Council's approach is reasonable as it explains different technologies and helps understand their impacts and risks; Parliament Draft, on the other hand, has a more inclusive wording. Therefore, a comprehensive article that includes the technologies of AI systems would better explain their properties, their impacts on individuals, and their risks to society.

Thus, one recommendation for the clarity of the definition of an AI system would be to:

> - Include technologies of AI systems explicitly in Article 3(1) AIA
>   AND
> - Explain the properties of the technologies in the recitals of AIA for further legal certainty.

5.2.2. Definition for Recommender Systems in Annex III

An important amendment to the AIA is the addition of recommender systems to high-risk AI systems in Annex III. This amendment ensures that online platforms classified as "very-large-online-platforms" ("**VLOPs**") under the Digital Services Act ("**DSA**") have risk management systems to monitor their AI system regularly.[234] However, the AIA does not include a definition for recommender systems.

The recommender system definition has evolved over the years,[235] although one comprehensive definition of recommender systems could include systems that guide individuals in a personalized way.[236] In the context of online platforms, this definition would include various practices like content curation, advertisements, and item recommendation, solving possible interpretational issues. While the best approach would be to incorporate a binding definition in the article, alternatively, providing explanations or examples in recitals would also help practitioners.

---

[233] Parliament Draft AIA, Article 3(1)(1).

[234] Parlimanent Draft AIA, Article 9.

[235] Burke R, Felfernig A, and Göker M H, 'Recommender Systems: An Overview' [2011] Association for the Advancement of Artificial Intelligence, <https://www.researchgate.net/publication/220604600_Recommender_Systems_An_Overview> accessed 28 May 2023.

[236] Alexander Felfernig and Robin Burke, 'Constraint-Based Recommender Systems: Technologies and Research Issues'(2008)<https://www.researchgate.net/publication/234797128_Constraint-based_recommender_systems_Technologies_and_research_issues > accessed 28 May 2023.

Therefore, one amendment could be to:

- • Include a detailed definition for "recommender systems" that underlines the *personalization* aspect in Article 3 AIA
  <u>OR</u>
- • Include explanations or examples in the recitals of AIA.

5.2.3.   The Classification of Online Platforms in Annex III

The DSA currently applies to seventeen VLOPs, including Amazon Store, Facebook, Instagram, LinkedIn, Pinterest, TikTok, Twitter, and Youtube,[237] and manipulative and discriminatory practice examples of these online platforms have been used in this thesis. Among the examples, Google was mentioned several times, although it is designated as a "very-large-online-search-engine" ("**VLOSEs**") under the DSA.[238] As a result, Google's recommender system is not considered as a high-risk AI system. This lack of recognition of VLOSEs, risks creating discrepancies in practice and hinders the overall protection of individuals in the digital realm of online platforms. To address this issue, both VLOPs and VLOSEs should be included within the scope of high-risk AI systems.

Therefore, one amendment could be:

- • 	The addition of VLOSEs classified under the DSA to Annex III, paragraph 1, point 8.

5.2.4.   Significant Harm Clarity to Article 5 AIA

Parliament Draft Article 5 AIA includes a significant harm requirement for prohibiting manipulative and discriminatory AI systems but lacks a clear definition or guidance on "significant harm." This ambiguity poses challenges in identifying manipulative and discriminatory AI systems since the concept of "significant harm" varies across different situations. Including a comprehensive definition or explanation of significant harm ensures consistent application of the AIA and enhances legal certainty, especially for digital environments where the harm is not directly visible.

The context and the nature of manipulation are important factors when determining the requirements of significant harm, as this prohibition under Article 5 AIA directly impacts *morally acceptable* practices on online platforms, such as content moderation and personalized recommendations. WP29's approach to interpreting the "similar significant effect" requirement

---

[237] 'DSA: Very Large Online Platforms and Search Engines' (*European Commission - European Commission*) <https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413> accessed 18 June 2023.
[238] ibid.

under Article 22(1) GDPR and their analysis of a context-based interpretation of targeted advertisements is a valuable guideline.

Therefore, another amendment that helps with the determination of acceptable manipulative practices is:

- Clarity over the "significant harm" concept under Article 5 AIA with a context-based interpretation.

5.2.5.   Feasibility of High-Risk AI System Obligations

High-risk AI systems are 'strictly' regulated under the AIA in line with the general principles, including establishing risk management systems and transparency requirements. However, there might be some feasibility problems due to the technology of AI systems. The requirements include identifying, estimating, and evaluating known and reasonably foreseeable risks and communicating with the relevant stakeholders if the measures do not eliminate significant risks.[239] While it is important to safeguard individuals, determining the foreseeable risks can be hard, especially with ML algorithms and their lack of explainability. Interestingly, unlike the datasets requirements, which have technical feasibility limitations within a specific market segment,[240] the EP does not include a limitation for this obligation. To ensure smooth enforcement, it is crucial to clarify the grounds of "reasonable" risks and limit the obligation to the technical feasibility of the AI system.

Thus, an amendment to help with the technical feasibility is:

- Incorporating clear grounds to determine "reasonable" risks to Article 9 AIA, AND
- Incorporating technical feasibility limitation to high-risk AI system obligations.

5.3.   Enforcement and Interpretation of Existing Rules

Given the potential risks of ADM, it is also important not to focus merely on specific provisions for ADM to find ways for amendments. Both legislations have several rules that can serve as protection over discrimination and manipulation of behavior arising from ADM systems in online platforms. Hence, it is crucial to consider two additional recommendations.

Firstly, increased enforcement of existing GDPR rules such as data protection impact assessments ("**DPIA**"),[241] and upholding general data processing principles[242] can contribute to the protection of individuals.

---

[239] Parliament Draft AIA, Article 9(4)(b)
[240] Parliament Draft AIA, Article 10(1)
[241] GDPR, Article 35.
[242] Article 29 Data Protection Working Party (n 24).

The GDPR requires DPIAs in specific situations, including when the technology systematically and extensively evaluates personal aspects based on automated processing, such as profiling.[243] Although DPIAs can become a hard task for the data controller to carry out under the complexity and opacity of the ML algorithms, they provide transparency and protection to individuals about manipulative and discriminatory practices by including the purposes, necessity, and proportionality of the processing, the assessment of risks, and risk mitigation measures. [244] Therefore, the DPIA obligation should be interpreted in a technically feasible manner for ADM and should be strictly enforced to assist transparency.

Enforcing general data protection principles set out in the GDPR is crucial for addressing discrimination and manipulation of behavior.[245] Fairness and transparency,[246] accuracy,[247] and accountability principles[248] also have particular importance. Infringement of compliance with these principles is subjected to administrative fines of up to 20.000.000 Euros, or up to 4% of total worldwide annual turnover.[249] The strong enforcement of these principles by the national data protection authorities may ensure data controllers adhere to the fundamental principles of data processing, considering the deterrent effect of the penalties.

Secondly, there is a need for a collective interpretation of the GDPR and the AIA for effective AI systems. This interpretation should consider the technological limitations, such as the opacity and the lack of explicability of the ADM and balance the technical restrictions of the AI system and the protection of individuals against discrimination and manipulation of behavior. Such collective interpretation is also beneficial for understanding the extent of manipulative and discriminatory AI systems prohibited under the AIA. This approach would also align with the new general principles in the Parliament Draft of AIA, which underlines the need of compliance with data protection rules in various stages of an AI system.[250]

Thus, two recommendations that could help are:

- Increased enforcement of DPIA and general data processing principles.
- Comprehensive interpretation of both the GDPR and AIA for a beneficial AI system.

5.4. Conclusion

This chapter focused on answering the fourth sub-question; "What could be the possible amendments to the current EU data protection regulation and proposed artificial intelligence regulation to address discrimination and manipulation of behavior arising from automated decision-making (ADM)?".

---

[243] GDPR, Article 35(1) and (3)(a).
[244] GDPR, Recital 75, Recital 90.
[245] Article 29 Data Protection Working Party (n 24).
[246] GDPR, Article 5(1)(a).
[247] GDPR, Article 5(1)(d).
[248] GDPR, Article 5(1)(e).
[249] GDPR, Article 83.
[250] Parliament Draft AIA, Article 4a.

In the GDPR, four amendments could help protect individuals against discrimination and manipulation of behavior arising from ADM. ***First***, clarifying the "prohibition" element and "legal or similar significant effect" in Article 22(1) GDPR and "explicit consent" in Article 22(2) GDPR would ensure legal certainty. ***Second***, enhancing Article 22(3) GDPR by adding minimum safeguards may enhance individual protection. ***Third***, incorporating clear definitions of "meaningful information about the logic involved" and "envisaged consequences of the processing" in Articles 13, 14, and 15 GDPR to balance technical limitations. ***Fourth***, there should be a clear distinction between *ex-ante* information provided in Articles 13-14 GDPR and *ex-post* information provided under Article 15 GDPR to incorporate a user-centric and situation-based transparency requirement.

In the AIA, five amendments could help; ***first***, the AI system definition should include an explanation of the underlying technology to acknowledge the possible risks of the technology. ***Second***, a recommender system definition should be added to the AIA for legal clarity. ***Third,*** VLOSEs should be included in high-risk AI systems of online platforms alongside VLOPs to have a complete representation of online platforms. ***Fourth,*** significant harm requirements for prohibited AI systems should be defined with clear grounds for legal certainty over morally acceptable practices on online platforms. ***Fifth***, risk management system requirements for high-risk AI systems should balance out technical feasibility and transparency better by incorporating reasonable limitations.

Apart from amendments, there are other solutions, i.e., increased enforcement of the existing general data processing articles and DPIA requirements for ADM and collective interpretation of the GDPR and the AIA articles together for a beneficial AI system integration.

6.   Conclusion

The main aim of this thesis was to answer the research question: "To what extent is the current EU legal framework for data protection and the proposed artificial intelligence regulation adequate to address discrimination and manipulation of behavior arising from automated decision-making (ADM) used in online platforms?"

The short answer to this question is that the current regulations address discrimination and manipulation of behavior arising from ADM and that there is no need for a new regulation.

Elaborating on this, ADM is a system that automates decisions about individuals that have rule-based or machine-learning algorithms. Online platforms use ADM in content curation, personalized recommender systems, and personalized, targeted advertisements to provide a more personalized user experience. While being efficient, they risk having adverse effects on the individuals, such as intentional and unintentional discrimination and manipulation of behavior. Although illegal, online platforms use ADM to intentionally discriminate against people based on *protected characteristics* generating unjust treatment of individuals. Besides illegal discriminatory practices, online platforms can use other grounds against individuals to exploit their vulnerabilities, leading to another problematic discriminatory practice. On the other hand, not all manipulative practices are problematic. Morally acceptable manipulative practices, such as personalized recommendations, are not problematic if they provide a personalized user experience without *ulterior motives*. Morally unacceptable manipulative practices, such as spreading disinformation through ADM to manipulate users or advertisements that exploit vulnerabilities, are more problematic.

The GDPR and the AIA have different regulatory approaches to protect individuals from discrimination and manipulation of behavior. While the GDPR regulates general data processing activities through ADM, the AIA focuses on the risks and impacts of AI systems. The GDPR prohibits the data processing activities of ADM unless one of the exemptions exists; however, it does not specifically address discrimination and manipulation of behavior arising from the use of ADM. Data subject rights provided in the GDPR play an important role in addressing the impacts of the ADM. Conversely, the AIA directly regulates manipulative and discriminatory AI systems with a full prohibition. While, initially, this approach "solves" manipulative practices on online platforms, when applied to online platforms will include the main activities of such platforms, including content moderation, personalized recommendations, and personalized advertisements that might not as problematic. AIA also acknowledges online platforms specifically and regulates recommender systems as high-risk AI systems with multifaceted requirements.

Both regulations currently have ambiguity problems and are open for amendments to ensure legal certainty and increase their adequacy in addressing problems of ADM. Individuals' protection levels will increase as the legal grounds become clearer in the regulations. Both regulations provide valuable rules that should be collectively interpreted to benefit ADM on online platforms. Once both regulations are in force and applicable throughout the EU, a clear interpretation of the GDPR and the AIA and strong enforcement of all rules will be necessary to ensure the protection provided in both regulations can be transposed into the practical world of online platforms.

Sources

I. Primary Sources

A. Treaties, regulations, legislations, etc.:

1) Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950, ETS 5, <https://www.refworld.org/docid/3ae6b3b04.html> accessed 30 March 2023

2) European Union, Charter of Fundamental Rights of the European Union, 26 October 2012, 2012/C 326/02, <https://www.refworld.org/docid/3ae6b3b70.html> accessed 30 March 2023

3) Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) And amending certain Union legislative acts COM(2021)206 final – 2021/0106(COD), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206> accessed 10 March 2023

4) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1, < https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679&qid=1688059883993> accessed 10 March 2023

5) Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) OJ L 277/1 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065&qid=1688059965544> accessed 10 June 2023

B. Case Law:

1) European Court of Human Rights, Biao v. Denmark (Grand Chamber), No.38590/10, 24 May 2016, para. 103

C. Guidelines, Studies, Opinions, etc.:

1) 'Guide on Article 14 on the European Convention on Human Rights and on Article 1 of Protocol No.12 to the Convention' (European Court of Human Rights 2022),<https://www.echr.coe.int/documents/d/echr/Guide_Art_14_Art_1_Protocol_12_ENG> accessed 10 June 2023

2) Article 29 Data Protection Working Party, 'Guidelines on Automated Decision-Making and Profiling for the Purposes of Regulation 2016/679' (2017) WP251rev.01 <https://ec.europa.eu/newsroom/article29/items/612053> accessed 7 March 2023

3) Article 29 Data Protection Working Party, 'Guidelines on Consent under Regulation 2016/679' (2017) WP259, <https://ec.europa.eu/newsroom/article29/items/623051/en> accessed 1 June 2023

4) Article 29 Data Protection Working Party, 'Statement on the Role of a Risk-Based Approach in Data Protection Legal Frameworks' (2014) WP 218, <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf> accessed 23 April 2023.

5) Council of EU, 'General Approach on Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' (Council of the European Union 2022) 2021/0106(COD) < https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf> accessed 1 June 2023.

6) European Data Protection Supervisor, 'Opinion on Online Manipulation and Personal Data' (European Data Protection Supervisor) Opinion 3/2018 <https://edps.europa.eu/data-protection/our-work/publications/opinions/online-manipulation-and-personal-data_en> accessed 1 June 2023.

7) European Parliamentary Research Service, 'Study on The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (European Parliamentary Research Service 2020) <https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530> accessed 1 June 2023.

II.   Secondary Sources

1) Angwin J, Scheiber N, Tobin A, 'Dozens of Companies Are Using Facebook to Exclude Older Workers From Job Ads' (ProPublica, 2017) <https://www.propublica.org/article/facebook-ads-age-discrimination-targeting> accessed 16 March 2023

2) Anonymous, 'A European Approach to Artificial Intelligence | Shaping Europe's Digital Future' (2023) <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> accessed 10 May 2023

3) Anonymous, 'AI & Machine Learning' (Center for Democracy and Technology) <https://cdt.org/ai-machine-learning/> accessed 7 March 2023

4) Anonymous, 'AI Act: A Step Closer to the First Rules on Artificial Intelligence | News | European Parliament' (2023) <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> accessed 16 May 2023

5) Anonymous, 'Amplifying Artist Input in Your Personalized Recommendations — Spotify' <https://newsroom.spotify.com/2020-11-02/amplifying-artist-input-in-your-personalized-recommendations/> accessed 21 June 2023

6) Anonymous, 'Antiquity' (A History of Artificial Intelligence) <https://ahistoryofai.com/antiquity/> accessed 7 March 2023

7) Anonymous, 'Are Facebook Job Ads Discriminatory? Company Accused of Bias against Women, Older Workers' (USA TODAY) <https://www.usatoday.com/story/money/2022/12/01/facebook-jobs-ads-discrimination-women-older-workers/10810589002/> accessed 22 March 2023

8) Anonymous, 'Classification: True vs. False and Positive vs. Negative | Machine Learning' (Google for Developers) <https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative> accessed 21 June 2023

9) Anonymous, 'Data Protection in the EU' (2021) <https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en> accessed 28 May 2023

10) Anonymous, 'Discrimination: What It Is and How to Cope' (https://www.apa.org) <https://www.apa.org/topics/racism-bias-discrimination/types-stress> accessed 7 March 2023

11) Anonymous, 'Discrimination: Your Rights' (gov.uk) <https://www.gov.uk/discrimination-your-rights> accessed 22 March 2023

12) Anonymous, 'DSA: Very Large Online Platforms and Search Engines' (European Commission - European Commission) <https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413> accessed 18 June 2023

13) Anonymous, 'EU's AI Law Needs Major Changes to Prevent Discrimination and Mass Surveillance' (European Digital Rights (EDRi)) <https://edri.org/our-work/eus-ai-law-needs-major-changes-to-prevent-discrimination-and-mass-surveillance/> accessed 4 June 2023

14) Anonymous, 'European Union Accession to the European Convention on Human Rights - Questions and Answers - Portal - www.coe.int' (Portal) <https://www.coe.int/en/web/portal/eu-accession-echr-questions-and-answers> accessed 21 June 2023

15) Anonymous, 'Facebook Adds "Feed Filter Bar" News Feed Control Tool, New Options to Restrict Post Comments' (Social Media Today) <https://www.socialmediatoday.com/news/facebook-adds-feed-filter-bar-news-feed-control-tool-new-options-to-rest/597664/> accessed 15 March 2023

16) Anonymous, 'New Ways to Customize Your Facebook Feed | Meta' <https://about.fb.com/news/2022/10/new-ways-to-customize-your-facebook-feed/> accessed 15 March 2023

17) Anonymous, 'How Google Photos Uses Machine Learning to Create Customized Albums' Christian Science Monitor <https://www.csmonitor.com/Technology/2016/0324/How-Google-Photos-uses-machine-learning-to-create-customized-albums> accessed 14 March 2023

18) Anonymous, 'IBM100 - Deep Blue' <https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/> accessed 7 March 2023

19) Anonymous, 'MEPs Ready to Negotiate First-Ever Rules for Safe and Transparent AI | News | European Parliament' (2023) <https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai> accessed 16 June 2023

20) Anonymous, 'Netflix Research - Recommendations' <https://research.netflix.com/research-area/recommendations> accessed 14 March 2023

21) Anonymous, 'Noise: The Other Side Of Human Error' <https://www.forbes.com/sites/hecparis/2021/10/29/noise-the-other-side-of-human-error/> accessed 21 June 2023

22) Anonymous, 'Poverty Inextricably Linked to Discrimination and Racism – UN Special Rapporteur' (OHCHR) <https://www.ohchr.org/en/press-releases/2013/11/poverty-inextricably-linked-discrimination-and-racism-un-special-rapporteur> accessed 22 March 2023

23) Anonymous, 'Ranking Results – How Google Search Works' (Google Search - Discover How Google Search Works) <https://www.google.com/search/howsearchworks/how-search-works/ranking-results/> accessed 14 March 2023

24) Anonymous, 'Recognizing People in Photos Through Private On-Device Machine Learning' (Apple Machine Learning Research) <https://machinelearning.apple.com/research/recognizing-people-photos> accessed 14 March 2023

25) Anonymous, 'Recommender System – Amazon Personalize – Amazon Web Services' <https://aws.amazon.com/personalize/> accessed 21 June 2023

26) Anonymous, 'Regulatory Framework Proposal on Artificial Intelligence | Shaping Europe's Digital Future' (2023) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 4 June 2023

27) Anonymous, 'Social Media Apps Are "deliberately" Addictive to Users - BBC News' <https://www.bbc.com/news/technology-44640959> accessed 5 June 2023

28) Anonymous, 'The History of Amazon's Recommendation Algorithm' (Amazon Science, 2019) <https://www.amazon.science/the-history-of-amazons-recommendation-algorithm> accessed 14 March 2023

29) Anonymous, 'The State of AI in 2022 - and a Half Decade in Review' <http://ceros.mckinsey.com/mckinsey-commentary-ai-meyhew-mobile-1-1> accessed 7 March 2023

30) Anonymous, 'Using Technology to More Consistently Apply Age Restrictions' (blog.youtube) <https://blog.youtube/news-and-events/using-technology-more-consistently-apply-age-restrictions/> accessed 14 March 2023

31) Anonymous, 'Vision, Mission & Values' (AlgorithmWatch) <https://algorithmwatch.org/en/vision-mission-values/> accessed 22 March 2023

32) Anonymous, 'What Are Customer Profiling and Its Undeniable Benefits to Business?' (Quantzig, 2021) <https://www.quantzig.com/blog/customer-profile/> accessed 21 June 2023

33) Anonymous, 'What Is Automated Individual Decision-Making and Profiling?' (2022) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/> accessed 7 March 2023

34) Anonymous, 'What Is Machine Learning and Why Is It Important?' (Enterprise AI) <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML> accessed 7 March 2023

35) Anonymous, 'When Does the Charter Apply?' <https://commission.europa.eu/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights/when-does-charter-apply_en> accessed 21 June 2023

36) Baeza-Yates R, 'Personalization, Bias and Privacy.' (2020) <https://dl.acm.org/doi/abs/10.1145/3386392.3399994?casa_token=EwJKdT2wj00A

AAAA:fKB9lQNaKGKg7bnVUcGemHAiI4sKaGVl7yB45WiJ2o0AALmxMFwbdECQtb
XB2Iq8gsFga8B48byrQw> accessed 1 July 2023

37) Beer D, 'Why Humans Will Never Understand AI' <https://www.bbc.com/future/article/20230405-why-ai-is-becoming-impossible-for-humans-to-understand> accessed 21 June 2023

38) Brkan M, 'Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond' (2019) 27 International Journal of Law and Information Technology 91, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3124901> accessed 28 March 2023

39) Brownlee J, 'Difference Between Algorithm and Model in Machine Learning' (MachineLearningMastery.com, 2020) <https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/> accessed 7 March 2023

40) Büchi M, Fosch-Villaronga M, Lutz C, Tamo-Larrieux A, Velidi S, and Viljoen S, 'Chilling Effects of Profiling Activities: Mapping the Issues' [2019] SSRN Electronic Journal <https://www.ssrn.com/abstract=3379275> accessed 20 March 2023

41) Burke R, Felfernig A, and Göker M H, 'Recommender Systems: An Overview' [2011] Association for the Advancement of Artificial Intelligence, < https://www.researchgate.net/publication/220604600_Recommender_Systems_An_ Overview> accessed 28 May 2023

42) Cadwalladr C and Graham-Harrison E, 'Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach' The Guardian (2018) <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election> accessed 19 March 2023

43) Castets-Renard C, 'Accountability of Algorithms in the GDPR and beyond: A European Legal Framework on Automated Decision-Making' [2019] SSRN Electronic Journal <https://www.ssrn.com/abstract=3391266> accessed 24 March 2023

44) Cormen T H, Introduction to Algorithms (3rd ed, MIT Press 2009)

45) Custers B, 'Profiling and Predictions: Challenges in Cybercrime Research Datafication', Lavorgna A and Holt T J, (eds), 'Researching Cybercrimes' (1st edn, 2021) < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4016165> accessed 7 March 2023

46) de Hert P, Lazcoz G, 'Radical Rewriting of Article 22 GDPR on Machine Decisions in the AI Era' (European Law Blog, 2021) <https://europeanlawblog.eu/2021/10/13/radical-rewriting-of-article-22-gdpr-on-machine-decisions-in-the-ai-era/> accessed 18 June 2023

47) Diakopoulos N, 'Algorithmic Accountability Reporting: On the Investigation of Black Boxes' (Columbia Journalism School, TOW Center for Digital Journalism 2013) <https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2> accessed 7 March 2023

48) Edwards L and Veale M, 'Slave to the Algorithm? Why a "right to an Explanation" Is Probably Not the Remedy You Are Looking For' (LawArXiv 2017) preprint <https://osf.io/97upg> accessed 21 June 2023

49) Felfernig A and Burke R, 'Constraint-Based Recommender Systems: Technologies and Research Issues' (2008)<https://www.researchgate.net/publication/234797128_Constraint-

based_recommender_systems_Technologies_and_research_issues> accessed 28 May 2023

50) Fry H, Hello World: How to Be Human in the Age of the Machine (2019)

51) Galli F, Algorithmic Marketing and EU Law on Unfair Commercial Practices, vol 50 (Springer International Publishing 2022) <https://link.springer.com/10.1007/978-3-031-13603-0> accessed 15 March 2023

52) Gayle D, 'Google Accused of Racism after Black Names Are 25% More Likely to Bring up Adverts for Criminal Records Checks' (Mail Online, 2013) <https://www.dailymail.co.uk/sciencetech/article-2273741/Google-accused-racism-black-names-25-likely-bring-adverts-criminal-records-checks.html> accessed 24 March 2023

53) Gellert R, 'The Role of the Risk-Based Approach in the General Data Protection Regulation and in the European Commission's Proposed Artificial Intelligence Act: Business as Usual?' (2021) 3, <https://jelt.padovauniversitypress.it/2021/2/2> accessed 28 May 2023

54) Ginani A, 'Why Social Media Platforms Need Your Personal Data' (Acquaint SoftTech, 2019) <https://medium.com/acquaint-softtech/why-social-media-platforms-need-your-personal-data-7d23dc115555> accessed 24 March 2023

55) Goodman B, 'Discrimination, Data Sanitisation and Auditing in the European Union's General Data Protection Regulation', European Data Protection Law Review, 495, (2nd Volume, 2016) <https://edpl.lexxion.eu/article/edpl/2016/4/8> accessed 7 March 2023

56) GreekDataGuy, 'Machine Learning Is Getting Easier. Software Engineering Is Still Hard' (Medium, 2020) <https://towardsdatascience.com/machine-learning-is-getting-easier-software-engineering-is-still-hard-d4e8320bc046> accessed 21 June 2023

57) Griffin R, 'Tackling Discrimination in Targeted Advertising: US regulators take very small steps in the right direction – but where is the EU?' [2022] Verfassungsblog <https://verfassungsblog.de/targeted-ad/> accessed 23 March 2023

58) Harris J G and Davenport T H, 'Automated Decision Making Comes of Age' (MIT Sloan Management Review 2005) <http://pubsonline.informs.org/do/10.1287/8943f842-86f8-4d42-9a64-9a7cd07b31f5/abs/> accessed 20 June 2023

59) Hawath M, 'Regulating Automated Decision-Making: An Analysis of Control over Processing and Additional Safeguards in Article 22 of the GDPR.' (2021) 7 European Data Protection Law Review 161, <https://edpl.lexxion.eu/article/EDPL/2021/2/6> accessed 20 March 2023

60) Heidary K, Custers B, Pluut H, and van der Rest J P, 'A Qualitative Investigation of Company Perspectives on Online Price Discrimination' (2022) 46 Computer Law & Security Review 105734 <https://doi.org/10.1016/j.clsr.2022.105734> accessed 24 March 2023

61) Hill K, 'How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did' (Forbes) <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/> accessed 14 March 2023

62) Hogg R, 'Elon Musk Called Twitter Newsfeed Algorithms "manipulative," Prompting a Reply from Founder Jack Dorsey, Who Disagreed with the Claim' (Business Insider) <https://www.businessinsider.com/musk-and-dorsey-argue-over-manipulative-twitter-algorithms-2022-5> accessed 19 March 2023

63) Jon Keegan, 'How Political Campaigns Use Your Phone's Location to Target You – The Markup' (2022) <https://themarkup.org/privacy/2022/11/08/how-political-campaigns-use-your-phones-location-to-target-you> accessed 18 June 2023

64) Jongepier F and Wieland J W, 'Microtargeting people as a mere means', Fleur Jongepier, Micheal Klenk (eds), The Philosophy of Online Manipulation (1st edn, Routledge, 2022) <https://www.taylorfrancis.com/books/9781003205425> accessed 19 March 2023

65) Jongepier F, Klenk M, 'The Philosophy of Online Manipulation', Jongepier F, Klenk M (eds), The Philosophy of Online Manipulation (1st edn, Routledge, 2022) <https://www.taylorfrancis.com/books/9781003205425> accessed 19 March 2023

66) Kaminski M E and Malgieri G, 'Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations' (2021) 11 International Data Privacy Law <https://academic.oup.com/idpl/article/11/2/125/6024963> accessed 28 May 2023

67) Keeling G and Burr C 'Digital Manipulation and Mental Integrity', Jongepier F, Klenk M (eds), The Philosophy of Online Manipulation (1st edn, Routledge, 2022) <https://www.taylorfrancis.com/books/9781003205425> accessed 19 March 2023

68) Klenk M, '(Online) Manipulation: Sometimes Hidden, Always Careless' (2022) 80 Review of Social Economy 85 <https://www.tandfonline.com/doi/full/10.1080/00346764.2021.1894350> accessed 7 March 2023

69) Llamas Covarrubias J Z, 'Algorithmic Transparency as a Foundation of Accountability' [2022] SSRN Electronic Journal <https://www.ssrn.com/abstract=4160223> accessed 3 June 2023

70) Logg J M, Minson J A and Moore D A, 'Algorithm Appreciation: People Prefer Algorithmic to Human Judgment' (2019) 151 Organizational Behavior and Human Decision Processes 90 < https://www.sciencedirect.com/science/article/abs/pii/S0749597818303388> accessed 7 March 2023

71) Machkovech S, 'Report: Facebook Helped Advertisers Target Teens Who Feel "Worthless" [Updated]' (Ars Technica, 2017) <https://arstechnica.com/information-technology/2017/05/facebook-helped-advertisers-target-teens-who-feel-worthless/> accessed 20 March 2023

72) Mahesh B, 'Machine Learning Algorithms - A Review' (2018) 9 < https://www.ijsr.net/get_count.php?paper_id=ART20203995> accessed 7 March 2023

73) Malgieri G and Comandé G, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 243 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3088976> accessed 28 May 2023

74) Mendoza I and Bygrave L A, 'The Right Not to Be Subject to Automated Decisions Based on Profiling' , Synodinou T E and others (eds), EU Internet Law (Springer International Publishing 2017) <http://link.springer.com/10.1007/978-3-319-64955-9_4> accessed 18 May 2023

75) Mittelstadt B D, Allo P, Taddeo M, Wachter S and Floridi L, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951716679679> accessed 7 March 2023

76) Nišević M, Sears A M, Fosch-Villaronga E, and Custers B 'Understanding the Legal Bases for Automated Decision-Making under the GDPR', Kosta E, Leenes R and Kamara I (eds), 'Research Handbook on EU Data Protection Law' (1st edn, 2022), <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331720> accessed 24 March 2023

77) Noble S U, Algorithms of Oppression: How Search Engines Reinforce Racism (New York University Press 2018)

78) Nys T, Engelen B 'Commercial Online Choice Architecture', Jongepier F, Klenk M (eds), The Philosophy of Online Manipulation (1st edn, Routledge, 2022) <https://www.taylorfrancis.com/books/9781003205425> accessed 19 March 2023

79) Pariser E, The Filter Bubble (Penguin Books 2011)

80) Prince A E R and Schwarcz D, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105 Iowa Law Review < https://ilr.law.uiowa.edu/print/volume-105-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data> accessed 7 March 2023

81) Q Y, 'Algorithms and Resistance: User Behaviors Towards Instagram's Algorithmic Feed' (Medium, 2019) <https://medium.com/@yangq.design/algorithms-and-resistance-user-behaviors-towards-instagrams-algorithmic-feed-c090780ee7a0> accessed 20 March 2023

82) Roth E, 'Facebook and Instagram Will Delete "Sensitive" Ad Targeting Groups Linked to Race, Politics' (The Verge, 2021) <https://www.theverge.com/2021/11/9/22773038/meta-detailed-targeting-ads-facebook-instagram-messenger> accessed 23 March 2023

83) Rubel A, Castro C, and Pham A, 'Algorithms and Autonomy: The Ethics of Automated Decision Systems' (1st edn, Cambridge University Press 2021) <https://philpapers.org/rec/RUBAAA-5> accessed 10 May 2023

84) Seaver N, 'Knowing Algorithms', Vertesi J and Ribes D (eds), DigitalSTS - A Field Guide for Science & Technology Studies (Princeton University Press 2019) <https://digitalsts.net/wp-content/uploads/2019/11/26_digitalSTS_Knowing-Algorithms.pdf> accessed 7 March 2023

85) Smith C, 'Study Reveals the Scary Amount of Personal Data Apps Can Collect without Your Permission' (*BGR*, 2021) <https://bgr.com/tech/location-tracking-personal-data-study/> accessed 18 June 2023

86) Susser D, Roessler B and Nissenbaum H F, 'Online Manipulation: Hidden Influences in a Digital World' [2018] SSRN Electronic Journal <https://www.ssrn.com/abstract=3306006> accessed 19 March 2023

87) Tufekci Z, 'Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency.' [2015] Colo. Tech. LJ 13 <https://heinonline.org/HOL/LandingPage?handle=hein.journals/jtelhtel13&div=18&id=&page=> accessed 20 June 2023

88) Turing A M, 'Computing Machinery and Intelligence' [1950] Computing Machinery and Intelligence 433 <https://academic.oup.com/mind/article/LIX/236/433/986238> accessed 7 March 2023

89) Veale M and Zuiderveen Borgesius F, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 4/2021 Computer Law Review International <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3896852> accessed 28 May 2023

90) Wachter S, Mittelstadt B D, and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' [2017] International Data Privacy Law <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469> accessed 28 May 2023

91) Wakabayashi D, 'Does Anyone Know What Paper Towels Should Cost?' The New York Times (2022) <https://www.nytimes.com/2022/02/26/technology/amazon-price-swings-shopping.html> accessed 24 March 2023

92) Waldman A E, 'Power, Process, and Automated Decision-Making' (2019) 88 Fordham Law Review <https://ir.lawnet.fordham.edu/flr/vol88/iss2/9/> accessed 7 March 2023

93) Wulf J, 'Guidebook on Automated Decision-Making Systems and Discrimination' (Algorithm Watch 2022) <https://algorithmwatch.org/en/wp-content/uploads/2022/06/AutoCheck-Guidebook_ADM_Discrimination_EN-AlgorithmWatch_June_2022.pdf> accessed 16 March 2023

94) Zuiderveen Borgesius F, 'Study on Discrimination, Artificial Intelligence and Algorithmic Decision-Making' (Council of Europe 2018), <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73> accessed 25 April 2023