Law, Governance and Technology Series 58

Henrique Sousa Antunes ·
Pedro Miguel Freitas · Arlindo L. Oliveira ·
Clara Martins Pereira · Elsa Vaz de Sequeira ·
Luís Barreto Xavier *Editors*

Multidisciplinary Perspectives on Artificial Intelligence and the Law

OPEN ACCESS

Springer

Law, Governance and Technology Series

Volume 58

The *Law, Governance and Technology Series* is intended to attract manuscripts arising from an interdisciplinary approach in law, artificial intelligence and information technologies. The idea is to bridge the gap between research in IT law and IT-applications for lawyers developing a unifying techno-legal perspective. The series will welcome proposals that have a fairly specific focus on problems or projects that will lead to innovative research charting the course for new interdisciplinary developments in law, legal theory, and law and society research as well as in computer technologies, artificial intelligence and cognitive sciences. In broad strokes, manuscripts for this series may be mainly located in the fields of the Internet law (data protection, intellectual property, Internet rights, etc.), Computational models of the legal contents and legal reasoning, Legal Information Retrieval, Electronic Data Discovery, Collaborative Tools (e.g. Online Dispute Resolution platforms), Metadata and XML Technologies (for Semantic Web Services), Technologies in Courtrooms and Judicial Offices (E-Court), Technologies for Governments and Administrations (E-Government), Legal Multimedia, and Legal Electronic Institutions (Multi-Agent Systems and Artificial Societies).

Henrique Sousa Antunes • Pedro Miguel Freitas •
Arlindo L. Oliveira • Clara Martins Pereira •
Elsa Vaz de Sequeira • Luís Barreto Xavier
Editors

Multidisciplinary Perspectives on Artificial Intelligence and the Law

Springer

CATÓLICA
RESEARCH CENTRE
FOR THE FUTURE OF LAW
LISBON

Plataforma
de Apoio à
Investigação
em Direito
na Católica

Editors
Henrique Sousa Antunes
Faculty of Law
Universidade Católica Portuguesa
Lisbon, Portugal

Arlindo L. Oliveira
Instituto Superior Técnico
University of Lisbon
Lisbon, Portugal

Elsa Vaz de Sequeira
Faculty of Law
Universidade Católica Portuguesa
Lisbon, Portugal

Pedro Miguel Freitas
Faculty of Law
Universidade Católica Portuguesa
Porto, Portugal

Clara Martins Pereira
Durham Law School
Durham, UK

Luís Barreto Xavier
Faculty of Law
Universidade Católica Portuguesa
Lisbon, Portugal

ISSN 2352-1902 ISSN 2352-1910 (electronic)
Law, Governance and Technology Series
ISBN 978-3-031-41263-9 ISBN 978-3-031-41264-6 (eBook)
https://doi.org/10.1007/978-3-031-41264-6

This work was supported by PAIDC - Plataforma de Apoio à Investigação em Direito na Católica

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

The Ethical and Legal Challenges of Recommender Systems Driven by Artificial Intelligence

Eduardo Magrani and Paula Guedes Fernandes da Silva

Abstract In a hyperconnected world, recommendation systems (RS) are one of the most widespread commercial applications of artificial intelligence (AI), initially mostly used for e-commerce, but already widely applied to different areas, for instance, content providers and social media platforms. Due to the current information overload, these systems are designed mainly to help individuals dealing with the infinity of options available, in addition to optimizing companies' profits by offering products and services that directly meet the needs of their customers. However, despite its benefits, RS based on AI may also create detrimental effects—sometimes unforeseen—for users and society, especially for vulnerable groups. Constant tracking of users, automated analysis of personal data to predict and infer behaviours, preferences, future actions and characteristic, the creation of behavioural profiles and the microtargeting for personalized recommendations may raise relevant ethical and legal issues, such as discriminatory outcomes, lack of transparency and explanation of algorithmic decisions that impact people's lives and unfair violations of privacy and data protection. This article aims to address these issues, through a multisectoral, multidisciplinary and human rights'-based approach, including contributions from the Law, ethics, technology, market, and society.

1 Introduction

Artificial Intelligence (AI) is constantly increasing its presence in our daily lives, shaping the way we access information, interact with connected devices, share personal information, and socially interact with others (Privacy International 2018,

E. Magrani
Berkman Klein Center for Internet & Society at Harvard University, Harvard Law School, Cambridge, MA, USA

P. G. F. da Silva (✉)
Catholic University of Portugal, School of Porto, Faculty of Law, Porto, Portugal

H. Sousa Antunes et al. (eds.), *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, Law, Governance and Technology Series 58,
https://doi.org/10.1007/978-3-031-41264-6_8

p. 4). Progressively, new products and services based on this technology are made available, for instance, through audio-visual recommendations; spam filtering in e-mails; personalized news feeds on social media; search results on search engines; virtual assistants and even suggestions on best routes on traffic apps.

Even though the term "artificial intelligence" has existed since the mid-1950s, the growing popularity of these systems is associated with the currently growing of data availability, cheaper processing infrastructure, technological advances, and greater connectivity (Bigonha 2018, p. 2). In a nutshell, AI may be considered a huge field of study, which reunites different technologies that combine data, algorithms and computational power (European Commission 2020c, p. 2), capable of behaving similarly to human intelligence to achieve specific objectives, usually the solution of a specific question (European Commission 2018, p. 1).

In the current state of the art, AI contributes to social and economic benefits in different fields by improving the prediction of results, optimizing operations and resource allocation and customizing service delivery, providing significant competitive advantages for the companies that dominate it (European Commission 2020b, p. 1). However, despite potentially beneficial to people and society, AI also raises new challenges.

Therefore, the rapid development and thoughtless application of technology establish the necessity to implement ethical principles and regulations for its use on the agenda, especially when we talk about machines with the ability to learn by itself, generating highly unpredictability results (even without human intervention) and great potential to harm fundamental rights.

The scale and reach of AI systems, the trend toward rapid and careless implementation, and the immediate impact they have on the lives of many people, may reinforce existing problems, besides the creation of new ones (Andersen 2018, p. 14). The threat posed by AI, then, does not assume the form of a super-intelligent robot that dominates humanity, but results from its daily use, as is the case with recommender systems, which will be specifically analysed in the following topic.

2 What are AI's Recommender Systems?

In a hyperconnected world, recommender systems (RS) are one of the most widespread commercial applications of AI, initially introduced for e-commerce, but already widely applied in other fields, such as content providers and social media platforms (Sahu and Singh 2019, p. 1).

Due to the current information overload, these systems are primarily designed to help individuals deal with the countless options available, as well as optimizing companies' profit generation by offering products and services that directly meet their customers' needs (Zhang et al. 2020, pp. 1–2). So, ideally, while RS create better user experiences, they also help providers fulfil their purpose of increasing the number of sales and clicks and, hence, profits, as well as increasing user engagement and satisfaction across different platforms (Tejeda-Lorente et al. 2018, p. 3).

Given its effectiveness, the use of RS already covers different domains, including streaming (Netflix and Spotify), news (CNN and Google News), dating (Tinder and Grindr), food (Ifood and UberEats), travel (Booking and AirBnB), social media (Facebook and LinkedIn), search engines (Google) and e-commerce (Amazon) (Paraschakis 2018, pp. 2–3). In the current big data era, the basic idea of recommender systems is to use the different data sources available to infer and predict the interests, tastes, and future behaviour of users to recommend personalized content, products, or services (Aggarwal 2016, p. 1).

Therefore, RSs are considered an algorithmic information filtering tool, capable of assisting users in their decision-making process, shaping online experiences by indicating items that are likely to please them (Mazeh and Shmueli 2020, p. 1). The prediction of the items' usefulness for a given user varies according to the recommendation algorithm model used (Zhang et al. 2020, p. 2). Currently, there are three main models:

1. content-based approach—recommendations are sent based on descriptions of items previously approved by the user, either through direct assessments or inferred behaviours (Jannach et al. 2010, p. 4);
2. collaborative filtering—process information on behaviours and opinions of a community to predict items of interest to the target user, as long as the group and individual profiles are similar (Jannach et al. 2010, p. 13); and
3. knowledge-based approach—instead of historical data, this model combines features submitted by the user with knowledge about a specific area, such as marketing or sales information. It is more used for more complex and less frequent situations, such as carrying out financial transactions or buying cars, apartments and luxury items (Aggarwal 2016, pp. 14–16).

In addition to the three main models, there are also hybrid systems, which combine the strengths of each of the previous models to create more effective systems, and systems that consider context, such as information about time, location, emotions, and social relationships (Jannach et al. 2010, p. 21; Aggarwal 2016, p. 8).

Regardless of the model, sending personalized recommendations requires building a user profile (profiling) (Kanoje et al. 2015, pp. 1–2)[1] that summarizes their preferences, tastes, frequent behaviours, and interests. This information can be extracted either implicitly, from the monitoring of the individual's behaviour online, or explicitly, when the user himself directly provides his data, such as filling out forms (Jannach et al. 2010, pp. 1–2; Paraschakis 2017, p. 211).

In summary, RS are essentially composed of three steps: (1) collection of personal data, directly or indirectly provided by users (input). In the latter case, they include, for example, click flows, browsing history, structural information of

[1] Briefly, the behavioural profile is a set of patterns used to concisely describe the user from their data, which are processed to infer their characteristics, future behaviours, tastes and interests. This process allows classifying them into profiles, used to recommend personalized items to better satisfy them.

visited web pages and purchase records, observed and inferred from the constant monitoring of the individual online; (2) data processing for the creation of the user profile, which can be represented by, for example, groups of terms or keywords; (3) targeting personalized content in the form of recommendations (output) (Nadee 2016, pp. 16–23).

There is no doubt that RS provide benefits in terms of organization, time optimization and improvement of the individual's online experience, by helping them search for content, services and products of interest. However, this technology may also generate negative—sometimes unanticipated—effects for users and society, especially vulnerable groups. Constant monitoring, automated analysis of personal data to predict and infer individual behaviours, preferences and characteristics, the creation of behavioural profiles and, finally, the sending of personalized recommendations may raise relevant ethical and legal questions, as it will be analysed in the next topics.

3 Ethical and Legal Challenges Associated with RS

The development, implementation and use of complex recommender systems may lead to significant ethical and legal problems. Concrete or potential damages and violations of fundamental rights are already a consequence of this technology, such as the lack of transparency and explanation of results (algorithmic opacity), reduction of individual autonomy, exposure of users to unjustified violations of privacy and data protection, unconscious manipulation of behaviours and discrimination (Milano et al. 2019, pp. 5–6).

In order to mitigate some of these threats and damages from AI in RS, it is necessary to introduce an ethical and regulatory debate on possible limitations applicable to this technology. In addition to binding legislation, ethical guidelines is a first step that must also be considered to minimize the risks associated with these systems and, simultaneously, maximize their benefits (Ekstrand and Ekstrand 2016, p. 16).

For some years, there have been a worldwide concern to define ethical limits for AI. A growing number of initiatives from different stakeholders define recommendations and guidelines for building ethical, trustworthy and human-centred AI. By 2020, at least 84 initiatives of AI ethical principles had been mapped, coming from public and private organizations, especially from Europe and United States (Jobin et al. 2019, p. 391; Hartmann et al. 2020, p. 6).

Although most documents set out a general ethical framework for AI, which focuses on protecting vulnerable people and dealing with asymmetries of information and power (Beil et al. 2019, p. 4), as RS are based on AI algorithms, these common basic principles can be directly applicable to them (Jobin et al. 2019, pp. 391–396). Among the principles most cited by these documents are transparency, justice, non-maleficence, accountability, privacy, beneficence, freedom, autonomy and trust.

Thus, the analysis of this technology through an ethical principle approach may be a relevant starting point to contrast how far RS's development and use are from an adequate implementation, where it acts more beneficially than harmful to society. In this regard, to reach such an analysis, the principles of beneficence and maleficence play an important role.

In line with the principle of beneficence, AI-driven technologies, such as recommender systems, should be developed to create an "AI for good". In other words, technology must promote well-being, dignity, common good and sustainability in all its phases and designs, in order to benefit people, society and the planet (Guszcza et al. 2020, p. 72). In this sense, these tools must promote human potential, creating new opportunities that increase individual self-determination, autonomy, human agency, social cohesion, and individual and collective capacities (Floridi et al. 2018, p. 690).

Beneficial AI initiatives must achieve physical and emotional well-being at individual and collective levels, such as improving health care, providing public benefits, expanding positive educational outcomes, and creating safer environments (Guszcza et al. 2020, pp. 72–74). Specifically regarding RS, this principle is not intended to undermine the great benefits produced by them, but to ensure that these technologies work in favour of human beings and not against them.

For example, a well-designed RS to help sick or unhealthy individuals presents a great opportunity to help people achieve a better quality of life in accordance with beneficence (Ekstrand and Ekstrand 2016, p. 2). Currently, initiatives in this direction already exist, such as wearables with gamification techniques and other behavioural interventions in the form of "nudges" created to encourage healthier behaviours (Guszcza et al. 2020, p. 73).

Besides that, based on the principle of non-maleficence, recommender systems must be designed not to harm human beings in any way, avoiding predictable, unforeseen or unintentional damages, such as biased recommendations, facilitation of the spread of misinformation and violation of privacy and data protection rules (Guszcza et al. 2020, p. 71). When it comes to non-maleficence, the main point is to prevent any type of damage, whether from the intention or malpractice of an individual or unforeseen technological behaviour (Floridi et al. 2018, p. 697).

Therefore, to prevent and avoid harmful RS, it is essential to understand technological limitations to manage potential risks (Guszcza et al. 2020, pp. 71–72). This principle emphasizes the alarming need to have AI systems in accordance with the standards and recommendations of data protection, privacy, cybersecurity and safeguarding all human rights by design and by default, in addition to an effective accountability system in case of misuse.

Thus, adjustments and harmonizing agreements between beneficence and non-maleficence are common, which requires the balance of RS benefits and risks in practice (Floridi et al. 2018, p. 697). For example, when companies prevent, through automated techniques, harmful content from being recommended to protect their users, although AI filtering has beneficial intentions, it can violate individual freedom and autonomy. Therefore, in practice, it is important to carefully consider

the possible ways in which systems could be misused or cause unintended damage to mitigate their adverse effects (Ekstrand and Ekstrand 2016, p. 2).

In this sense, the principles of beneficence and non-maleficence, together with other ethical guidelines, connect and unfold in many different legal implications. Below, we highlight some of the main ethical and legal challenges that arise from the lens of these two values:

3.1 Opacity

Some AI experts compare this technology to a black box, as its processes and mode of operation would be beyond human capacity to understand (Floridi et al. 2018, p. 692), especially for people outside the field of technological study. This presumption is even more intense in the case of AI algorithms that interact in an open social environment and learn by interacting with the space in which they operate, when their automated decisions are difficult to explain even for experts. This frequent lack of transparency and explanation about the processes and values involved in the recommendation tools hinder the creation of better systems, that is, adequate to fundamental rights, ethical principles and centred on human beings (Milano et al. 2019, p. 16).

3.2 Discriminatory Bias

RS are created by people, which makes it susceptible to biased results. This consequence may arise as a result of the selected training data or (implicit) values held by technology developers, which may exacerbate systematic social discrimination, even unintentionally (European Parliament 2020, p. 15).

Due to the data-driven nature of the AI techniques used in the recommender system, the selection of the dataset for training must be well defined, otherwise it can be an important source of discrimination (Beil et al. 2019, p. 4). For example, when available data do not reflect the social diversity present in society, this population imbalance within the datasets is likely to generate bias against specific groups. In addition, biased content may also arise from feedback loops produced by the system for certain user groups, often reinforcing the racial and gender discrimination that already exists in society (Milano et al. 2019, pp. 12–13).

Within the processes performed by RS, profiling is one of the most likely to cause discrimination. With personal data, RS providers create profiles of their users as a parameter of aspects of their personality and interests, in order to label individuals according to certain patterns of habits, behaviours and tastes, which has great discriminatory potential, especially in the case of sensitive personal data (Mulholland and Frajhof 2019, pp. 269–270).

3.3 Privacy and Data Protection Violations

RS based on AI collect, analyse, and process a large amount of personal data. Thus, concerns about privacy and data protection grow as their use becomes commonplace and applicable in different areas, including in domains with highly privacy risks, such as healthcare and banking (Zhang et al. 2020, p. 14).

In this case, privacy-related risks may arise from all steps of the processing of user data. Considering the General Data Protection Regulation (Regulation 2016/679—GDPR) as a model, when data are collected by the algorithms of recommender systems and eventually shared with third parties—often without the implementation of security measures, valid consent (or other legal basis) and the provision of sufficient information to users—their privacy and personal data are violated, which is worsen in the case of data leakage and breach of anonymity (Milano et al. 2019, p. 7).[2]

In addition, RS' data processing may result in inferences and predictions of confidential and personal information, such as emotional states. Consequently, these systems can reach sensitive personal data (such as information about racial or ethnic origin, religious conviction, political opinion, health or sex life), from inferences extracted from personal data by automated processing for profiling and for the creation of personalized recommendations. Thus, significant privacy challenges are generated in this scenario, in addition to possible discrimination results (Privacy International 2018, p. 18).

3.4 Diminished Human Autonomy and Self-Determination

RS involves decision-making processes about their users and their contexts through the creation of behavioural profiles. This technology, capable of knowing potential users' preferences and adapting according to their presumed interests, raises important questions about privacy, autonomy and the ethics behind the adaptation processes (Privacy International 2018, p. 19).

Individual autonomy involves the capacity for free self-determination and the right to make choices based on personal beliefs, information, and values. For this, it is essential that the individual has a real and significant opportunity to make their own choices, properly informed and free from coercion, restrictions, or external influences, excessive or undue (Bernal 2014, pp. 24–25).

Thus, human autonomy is directly affected by RS, as they limit individual freedom, due to their control over influences that are transmitted to users in the form of recommendation, besides the fact that, when consent is used as a legal basis

[2] *See* also, on the GDPR, I.1—A Oliveira and M A T Figueiredo—Artificial intelligence: historical context and state of the art.

for personal data processing, it is rarely informed for the user, but used as an implicit condition for accessing a certain desired service (Varshney 2020, pp. 1–2).

In this sense, RS interferes with people's autonomy in the form of recommendation of all types of content, from music and movie to job opportunities, pushing users in a certain direction, generally related to their preferences drawn from their profiles, in an attempt to addict them to some types of content or limit the range of options to which they are exposed (Milano et al. 2019, p. 10). Some of these technologies act almost like traps to keep users engaged and connected to their platforms (Seaver 2018, p. 1), which allows greater availability of data to be collected and processed.

Moreover, the algorithmic profile of recommendation platforms also has a great impact on people's autonomy, as it can interfere with the experience of personal identity. First, systems based on user feedback (for example, collaborative filtering) do not create a specific and unique profile, but a collective one. Furthermore, classification is done by algorithms that analyse and infer tastes and preferences, which may not correspond to the appropriate social characteristics or categories with which the user identifies (Milano et al. 2019, p. 10). As mentioned before, the problem is also aggravated in the usual context of algorithms lack of explainability or transparency related to the creation of these profiles.

Thus, the use of recommender systems by bigtechs today, especially in social media, streaming and e-commerce, may also pose intentional risks to users' autonomy. According to their commercial interests, RS providers may also impose hidden influences on their users' behaviour, which is done through monitoring, behavioural tracking and exploitation of vulnerabilities and personal data for profiles creation, which are used to micro-targeting of content in the form of recommendations (Susser et al. 2019, p. 6). This process often occurs without the knowledge of the common user, which can interfere with their ability to self-determine and make truly autonomous choices (Susser et al. 2019, p. 13).

3.5 Polarization and Manipulation of Democratic Processes

Recommender systems and social media filters, by the nature of their design, take the risk of isolating users from exposure to different viewpoints. Even when the system correctly labels individuals, the effects produced by personalization may produce individual and collective harm by creating or exacerbating filter-bubbles[3]

[3] The idea of "filter bubble" was created by Eli Pariser to designate the phenomenon of algorithmic filtering of information, carried out on digital platforms such as social media and search engines, responsible for customizing the content that each user has access to, according to their interests, which causes the individual to be trapped in a "bubble" of information with which he agrees, while what he dislikes, shocks or disagrees with is hidden.

(Pariser 2011; Magrani 2014, pp. 118–119) and echo chambers (Sunstein 2007, pp. 43, 60, 217–218; Milano et al. 2019, pp. 13–14).[4]

Contents recommended on digital platforms, limited by these phenomena, represent high risks to public debate and the democratic process, as they may reinforce discriminatory biases and individual prejudices, increasing the susceptibility to polarization, hate speech and manipulation of public speech. As demonstrated by the Cambridge Analytica scandal, RS of streaming platforms and social media may become a place for sending targeted political propaganda (Milano et al. 2019, pp. 13–14).

Today, due to information overload, there's no doubt that recommender systems may mitigate this problem and help people manage their time efficiently. However, in this scenario, as much as technological recommendations can benefit users (helping individual performance in the process of choice, improving and diversifying decision making), they are also potentially questionable, as they influence people in a specific direction, and generate individual and social harm, such as information segregation, bubbles and behaviour manipulation. Thus, to ensure harmony with the principles of beneficence and non-maleficence, the system must be well designed not only to improve people's lives, but also to maintain full and effective control over themselves (Milano et al. 2019, p. 10), while avoiding harm and limiting risk.

4 Recommender Systems: Legal and Regulatory Challenges

Considering the growing importance of RS for our daily lives, simultaneously with the increase in their adverse effects, there is a huge need for action. Legal regulation initiatives must consider not only official ethical guidelines, but also the effective protection of human rights, starting from the basic premise that AI systems must work to do good, avoiding harm, not causing it.

Thus, as RS require the processing of personal data, the issues arising from these technologies have been addressed by data protection rules worldwide (Bioni and Luciano 2019, p. 2). In this scenario, the European Union (EU) GDPR plays an important role as a regulatory model that has inspired many others around the world (Silva 2020, p. 214), phenomenon known as the Brussels Effect.[5] Although the regulation does not specifically address RS or AI itself, it does address their fundamental processes, such as the processing of personal data for automated decision-making, profiles creation and the recommendation of personalized content.

[4] The term "echo chamber" is used by Cass Sunstein to designate an environment in which individuals only find ideas, beliefs and opinions that coincide with their own, which reinforces their views and does not consider alternative ones. For him, this phenomenon can lead to fragmentation and polarization, being a threat to democracy.

[5] The term "Brussels Effect" was coined in 2012 by Professor Anu Bradford of the Columbia Law School (Bradford 2012).

Hence, to protect fundamental rights, guarantee informational self-determination and the free development of personality, the GDPR brings a series of obligations imposed on controllers and processors, which include a list of principles (art. 5), rights of the data subjects (chapter III) and legal basis for processing of personal data (articles 6 and 9). Thus, RS' platforms must adapt to these rules to protect personal data of individuals and, consequently, other human rights potentially threatened by RS (Human Rights Watch 2018).

First, RS' providers need to ensure that all activities with personal data (automated or not) comply with the principles, especially the obligation of a lawfully, fairly and transparent processing (lawfulness, fairness and transparency) and the definition of a specified, explicit and legitimate purposes, in accordance with the legal bases of articles 6 and 9 (purpose limitation). Also, data must be limited to what is strictly necessary to achieve this purpose (data minimization) and kept only for the necessary period for it (storage limitation). Finally, the process must guarantee data accuracy and quality, compliance with security standards (integrity and confidentiality), besides ensuring accountability that enables eventual liability for damages.

Along with the adequacy to the principles, to be considered lawful, the processing of personal data by RS must occur in accordance with one of the situations described in art. 6. At this point, it is important to mention that GDPR, as a rule, prohibits the processing of special categories of data in art. 9 and fully automated decision-making with detrimental effects on the data subject in art. 22, except in specific situations listed in both articles. For the last, exceptions include obtaining the explicit consent of the data subject; when it is necessary for entering into or the performance of a contract; or is authorized by Union or Member-State law (WP29 2017, pp. 34–35).

Besides, RS providers need to ensure, throughout data process, an effective and facilitated exercise of data subjects' rights, which are considered a logical outcome of the principles (WP29 2014, pp. 16–17). For example, as a consequence of the legal and ethical principle of transparency, the right to information (articles 13 and 14) stipulates that users must be kept informed and aware of the possible risks associated with data processing carried out by RS. With that, users may not limit themselves to short-term gains obtained with these systems that could, slowly, undermine their fundamental rights, such as autonomy, freedom and privacy.

Thus, it is the duty of providers to proactively inform, even without request, about rights, the existence of data processing and other related information, including clear, meaningful and understandable purposes and explanations on the functioning of RS algorithmic techniques, in particular the definition of profiles (WP29 2014, pp. 16–17; Tejeda-Lorente et al. 2018, p. 6). Furthermore, this information, when not actively disclosed, must be provided to the subjects upon request for access, according to art. 15 and recital 63.

When analysed together, information and access rights are considered powerful tools for individuals to exercise greater control over their data related to RS, as it allows them to have larger awareness and knowledge about the processes involved in sending personalized recommendations, allowing better decision-making that could

protect their rights (Van Ooijen and Vrabec 2018, p. 94). Also, with the information received or requested, users can exercise other rights of GDPR, such as rectification (art. 16), erasure (art. 17), restriction of processing (art. 18), portability (art. 20), object (art. 21, when possible) and contest fully automated decisions (art. 22.3). This ensures users' greater autonomy and control, preventing harmful and biased recommendations.

That said, as the automated creation of profiles and the sending of personalized recommendations based on these profiles are steps of RS, article 22 is a key element, as it permits automated decision-making, including profiling, that produces legal effects on data subjects, only in the specific hypothesis authorized by the regulation, such as when based on data subject's explicit consent. In this context, the individual has the right to obtain human intervention, express his or her point of view and to contest the automated decision of the RS. Still, considering the risks involved, GDPR creates for controllers the obligation to adopt safeguard measures to protect data subjects' rights, freedoms and interests, which may include, privacy by design techniques (art. 25) and the carrying out of data protection impact assessment (art. 35).

Furthermore, as these systems rely on algorithmic probability and often machine learning models to send recommendations, it is essential to grant the data subject the right to clear and adequate explanation of the fully automated decisions involving their data. This right to explanation may be extracted from the interpretation of articles 13, 14 and 22, together with recital 71 and the principle of transparency, creating a controller's obligation to significantly inform about the logic involved in all the automated processes until the effective decision making. Such explanation does not necessarily involve the complete opening of the algorithms, but just enough for the user to understand the reasons underlying the decision that affects him (WP29 2018, p. 25), which guarantees the exercise of other rights of GDPR, besides the protection of other human rights (Monteiro 2018, pp. 12–13).

Thus, within the scope of RS, the application of art. 22 and the right to explanation is essential to minimize the risks of the increasing use of algorithms to classify people into behavioural profiles (Silva 2020, p. 210), based on inference analyses and predictions about their characteristics, tastes, behaviours and interests, and then send personalized recommendations potentially harmful to users, which silently interfere with their autonomy, manipulate their decisions and violate guarantees of non-discrimination and privacy.

That said, there is no doubt that the GDPR creates a favourable background for data protection in the EU, becoming a worldwide inspiration, applicable to AI tools, including recommender systems, imposing significant obligations and requirements on data controllers (Bernal 2014, p. 14). Though, besides protecting and defending fundamental rights, according to art. 1, the regulation also produces positive effects for companies and governments, as its application prevents violations of rights and, thus, sanctions' imposition, helping in the use and development of technologies that are beneficial to society. For example, the right to challenge automated decisions allows RS users to contest inaccurate or discriminatory recommendations, as well as an opportunity for the provider to revise their system (Souza et al. 2021, p. 476).

However, with big data, growing importance of digital platforms and the rapid expansion of AI techniques, despite the regulation trying to improve the context of data protection and, hence, human rights, there is still a lot to be done. Some of its rules are still difficult or not convenient for RS providers to comply with, especially those related to AI techniques for profiling and automated decisions. In this context, RS providers may face difficulties in ensuring compliance with principles and rights in practice, due to technical opacity or trade secret rules, for instance. Yet, there are many open questions concerning the interpretation of legal provisions, especially regarding the rights of data subjects, such as the right to contest automated decisions and explanation.

4.1 Lack of Transparency

Although ethical principles and legal rules demand the transparency of AI systems, some of their uses may be opaque for individuals, regulators and even for their designers, which makes it difficult to challenge results. So, RS may have three distinct sources of opacity: (1) intentional opacity, usually associated with trade secret; (2) opacity as technical illiteracy; and (3) opacity that arises from the design and characteristics of the system, especially in the case of machine learning (Privacy International 2018, p. 26).

This absence or lack of transparency in RS makes it difficult to question the political, economic and cultural agendas that exist behind the personalized recommendations sent to each user of the platform, in addition to hiding possible algorithmic discriminations and silent manipulation of behaviours. Besides the potential for damaging fundamental rights, opacity hampers the detection and correction of biased data, invalid assumptions and flawed models (Paraschakis 2017, p. 214).

4.2 Trade Secret

Information about the functionality of RS algorithms is often intentionally poorly accessible to the public (Mittelstadt et al. 2016, p. 6). Software, algorithms and data involved in recommender systems applications are considered proprietary assets with high added value, being essential to maintain an organization's position in the competitive market (European Parliament 2020, p. 33).

Consequently, most companies and providers of these systems are still reluctant and refusing to disclose information related to the functioning of AI because of trade secret (Milano et al. 2019, p. 2), which leads to an intentional opacity of RS. In particular, the lack of transparent business models and practices represents a significant barrier to detecting cases of human rights violations, such as discriminatory recommendations and inferences (Wachter 2020, p. 2).

4.3 Constantly Changing Technology

The current state of technological development of the AI, which bases the RS, does not clarify what the next big evolution will be and what kind of use and levels of understanding of the technology we will be able to make in the future (European Parliament 2019, p. 8), which hamper the imposition of damage prevention obligations to organizations that use AI. Furthermore, the "black box" mentality, whereby AI systems are beyond human comprehension, still limits human's control over technology (Floridi et al. 2018, p. 692).

4.4 Difficulties of Implementation of Data Subjects' Rights in Practice

As a rule, a typical RS system work as a black box, as the final recommendation (output) is the only part available to the user (Paraschakis 2017, p. 214). Whatever the reason for creating opaque RS, this lack of transparency is an obstacle to the fulfilment of the right to explanation of GDPR, which also hinders human control over how data is treated and the exercise of other rights.

Furthermore, currently, there is an imbalance of decision-making power and knowledge in favour of RS providers and to the detriment of users. This informational asymmetry, driven by the opacity of AI systems, is also reinforced by the absence or poor understanding of individuals regarding their rights and how the technology works in practice (Mittelstadt et al. 2016, p. 6), that is, how the algorithms and data processing techniques act when predict and infer behaviours, create profiles and send personalized content.

When the logic behind recommender systems is not understandable to the user, the control and autonomy of the human being are disrespected. Therefore, when RS provider relies on consent for the processing of data, this consent is not, in fact, freely given, specific, informed and unambiguous, as the user does not have sufficient information and appropriate means to assess the risks involved in processing data that adheres (Mittelstadt et al. 2016, p. 7).

In addition, given the concern of companies to implement data protection rules that require essential information and explanation disclosure, individuals face an overload of consent requests, usually through extensive and complex privacy policies and cookie notification (Van Ooijen and Vrabec 2018, p. 94). Considering the limits of human rationality and lack of time, the user's evaluation and effective control are impaired, which ends up in the failure to make informed decisions (Bioni 2019).

Also, despite living in the era of hyperconnectivity, most people still have little technical knowledge, access to digital education and minimal understanding of data processing processes (Bioni 2019), making it even more difficult to make informed decision-making in the context of RS, especially when based on consent. In practice,

the consent incorporated in most RS providers' privacy policies neither empowers users nor guarantees the effective exercise of rights and their informational self-determination, functioning as an apparent legitimacy of the business models to the GDPR rules (Bittencourt and Gomes 2019, pp. 26–33).

Therefore, individuals are placed in a situation of informational, technical and economic asymmetry (Edwards and Veale 2018). Although data protection rules aim to protect fundamental rights by establishing rights of data subjects, there is still a lack of effectiveness in different situations, for example, when it comes to inferential data analysis using AI techniques.

With the current legal context, data subjects lack sufficient control and information about how their data is being used by RS to make inferences, predictions and assumptions about them. Thus, individuals face obstacles to exercising their data protection rights, especially explanation and challenge of automated decisions, which is even harder when confronted with the interests of controllers related to intellectual property and trade secret (Wachter and Mittelstadt 2019, pp. 5–6).

Hence, specifically regarding the rights of explanation and automated decision challenge, there are still many open questions, as its parameters are still under discussion. Given this uncertainty, the recognition of the right to explanation in practice is impaired, which also makes it difficult to exercise other rights, especially contesting and review automated decisions, since the user must access information about automated decision, and the RS itself, to gather conditions to expose how his or her data should be process and eventually find errors, discrepancies and erroneous correlations to be solved (Souza et al. 2021, p. 473).

4.5 Difficulties of Rules' Application

Some specific characteristics of RS, such as opacity (black box effect), can make it difficult to apply and verify compliance with ethical guidelines and legal rules, especially those arising from the GDPR. Due to their high complexity, unpredictability and autonomous behaviour, authorities and people affected by these systems may not have specific means to verify how a particular personalized recommendation was achieved and, thus, whether these rules were complied with (European Commission 2020c, pp. 10–12).

The current regulatory debate emphasizes the role of data protection in establishing the rights of data subjects, legal basis and principles, focusing on the role of accountability, which highlights the ethical principle of non-maleficence. For example, Article 58 (2) of the GDPR establishes supervisory authorities' corrective powers, such as the imposition of fines, to be applied according to the circumstances of each case, always in an effective, proportionate and dissuasive manner.

Considering RS, digital platforms should ensure that their content and activities respect human rights, especially data protection, privacy and equality, and are not susceptible to external attacks. An interesting point is that some challenges related to these systems are more difficult to address using only technological solutions,

requiring a more qualitative analysis based on the social context in which they operate (Milano et al. 2019, p. 16).

In this case, the application of the GDPR by the authorities must seek a fair balance between the rules of the law and technological advances, preventing companies from suffering from regulations that burden them excessively with administrative requirements and unrealistic data protection standards. The open question is whether States will enforce this measure without burdening corporations or impeding technological innovation.

4.6 Beyond Damage Prevention

The current RS regulation for data protection in the GDPR focuses on measures to prevent damage and ensure accountability in the event of its occurrence, in accordance with the AI's non-maleficence idea. However, technologies must also be regulated through beneficence, which enables the maximization of benefits for individuals and society.

Given the undoubted potential of AI, mainly through recommender systems, it is worth regulating it so that its benefits are increased, avoiding potential pitfalls. In due course, AI regulation also needs to focus research not only on making the technology more capable and accurate, but also on maximizing its societal benefits (Russell et al. 2015, p. 106), which may be accomplished throught prior human rights' assessmentns.

5 Strategies and Possible Solutions to the Challenges Created by RS

Currently, GDPR represents a strong system of fundamental rights' protection in the context of AI and automated decisions. In addition to establishing relevant principles, such as legality, data minimization, transparency, security, fairness and accountability, it also stipulates a series of rights that strengthen the user's control over their data and establishes obligations for those responsible for processing such data, which includes the publication of information, transparency and implementation of security measures (Souza et al. 2021, pp. 470–471).

However, given the progressive and constant complexity of recommender systems based on AI, regulation solely by data protection law is no longer sufficient. So, there are other ways to address the problems associated with RS, which also includes specific legal rules related to AI and business models that use it, besides other strategies beyond law, such as social norms, market initiatives and the ways systems' architecture (code) are developed.

5.1 Best Practices Beyond Law

In this scenario, all stakeholders related to RS must pay attention to ethical standards applicable to AI algorithms. As stated, there is a wide debate around these ethical guidelines that should guide the entire lifecycle of AI-based RS, including their development, implementation, and effective use. There is an urgent need for these tools to focus on human beings, protecting their interests and fundamental rights, in order to benefit the entire society (Beil et al. 2019, p. 1). Given the relevance of ethical parameters, such as transparency, accountability, non-discrimination, precaution, privacy and security, many of them have already been incorporated in regulations, as happened in GDPR principles, rules and rights.

That said, as recommender systems are embedded by autonomous and intelligent algorithms, creating legal and ethical issues, initiatives from multidisciplinary areas of expertise, such as data scientists, lawyers, legal research experts, social scientists and ethics experts are required (Currie et al. 2020, p. 752). In this sense, AI solutions must be developed and implemented through an intersectoral and multidisciplinary teams with the goal of optimizing their results towards ethics and legality (European Parliament 2020, p. 52).

5.1.1 Regulation by Technology: Strategies by Design and by Default

In the context of these "new" technologies that actively interfere in our daily lives, recommending personalized content and making automated decisions about us, ethics and human rights play an important role in their application in favour of the public good. Thus, RS regulation must also involve the design of the tool itself, aligned with ethical guidelines and the human rights from the beginning, as a central element of the systems architecture (Magrani et al. 2019, p. 128).

This "value-sensitive design" approach, including privacy, security, ethics and human rights (Magrani 2019, p. 235), suits the idea that the benefits and positive effects of AI should not only be guaranteed by compliance with the regulatory framework, but also ensured by default (Cavoukian 2009, p. 1), from the beginning of the development of the system and reinforced during its use, according to strategies by design and by default.

Consequently, ethical and legal principles, based on human rights and values, should serve as design criteria for the development of innovative uses of AI and also for the review of existing ones, in order to place the human being at the centre of the creation of RS models, guiding their implementation and use (Guszcza et al. 2020, p. 80), in accordance with what is already provided by art. 25 of GDPR.

Thus, in the short term, design can play a crucial role in addressing ethical and legal issues potentially triggered by RS. For instance, pop-up messages alerting users about the results of recommendations that consider their behavioural profile help to raise public awareness and exercise of rights. However, in the long term, it is essential that RS infrastructure apply by default ethical norms and principles,

such as transparency, non-discrimination, and justice, in all phases of the system (European Parliament 2020, p. 30).

5.1.2 Implementation of (Human Rights) Impact Assessments

Considering the high risks for users and society created by the recommender systems, which include manipulation, violation of privacy and data protection, discrimination and reduction of individual autonomy, the prior carry out of human rights impact assessment and evaluation of compliance with legislation and ethical guidelines are fundamental for RS to be used (European Commission 2020c, p. 23). Currently, however, these systems are still being implemented to the public without proper ethical, legal, and technical evaluation that can assess the possible impacts and risks associated with this technology in practice, which puts the rights of individuals at stake (Reisman et al. 2018, p. 4).

As much as art. 35 of the GDPR determines to carry out personal data protection impact assessments in some specific cases, it is understood as good practice that RS providers carry out assessments and audits on all automated AI decisions, including profiling, which may be done by testing, inspection, or certifications (European Commission 2020c, p. 23). Therefore, it is recommended to implement algorithm audits and algorithmic impact assessments so that the risks associated with these tools may be mapped, prevented and mitigated (Ada Lovelace Institute and DataKind UK 2020, p. 23).

In this sense, the algorithm audit in RS must assess both the data and the algorithms to look for possible biases (bias audit), in addition to assessing the level of adequacy of the system to existing legal regulations and ethical guidelines (regulatory inspection), especially in terms of human rights. In addition, vendors must also implement algorithmic impact assessment, including risk and impact assessment of algorithms, which may end up evaluating potential social impacts of recommender systems before and during their implementation in practice (Ada Lovelace Institute and DataKind UK 2020, p. 3).

Furthermore, such processes must be developed before and during the technology's interaction with users (Ada Lovelace Institute and DataKind UK 2020, p. 3). If the recommendation system is not approved in such assessments, failing to comply with legal and ethical requirements, identified failures must be solved or mitigated, through new tests or imposition of safeguards and safety mechanisms (European Commission 2020c, p. 23).

In addition to the prior control carried out by the recommendation providers themselves, it is important that a subsequent control is also carried out, not only through technology assessments, but also through documentation verification and even external audits by specialized organizations. Such compliance monitoring should be part of an ongoing market supervision framework for these technologies (European Commission 2020c, p. 23).

5.1.3 Guarantee of Greater Transparency and Explanation of AI (Explainable AI)

RS should be designed to explain its reasoning and allow humans to interpret results (recommendations). As previously mentioned, the explanation of functions and processes is vital to ensure the exercise of rights, transparency and accountability, which is in line with the legal interpretation of GDPR that established the right to explanation.

The explanation of recommender systems and their decisions, as a dimension of the principle of transparency, would enable greater balance between economic and social interests by allowing the existence of automated decisions and, simultaneously, reducing informational asymmetries between those responsible for data processing and the users of the system, as it makes the disclosure of information a legal obligation (Souza et al. 2021, p. 472).

According to the European Commission, the opacity of AI systems can be mitigated through transparency obligations (European Commission 2020c, p. 15), which include accessibility and understandability of information (Mittelstadt et al. 2016, p. 6). Without proper transparency in processes and decisions, in addition to concrete mechanisms that ensure clarification and effective information, users may have difficulties understanding the systems they use and their recommendations, which would make harder to ensure accountability in case of damage. Thus, explainable recommendation techniques are an essential approach to improve transparency, effectiveness, reliability and user satisfaction with systems (Zhang and Chen 2020, p. 77).

Explainable recommendations, for example, are essential for e-commerce, as they increase the persuasiveness of suggestions and, at the same time, help consumers to make efficient and informed online decisions. This strategy would facilitate the process of making AI technologies socially responsible by ensuring both commercial profits and benefits to users. In addition, some RS can provide essential and crucial information for sensitive decision-making, such as in medical treatment processes, where the explanation of recommended results is vital to ensure the effective safeguarding of other people's lives and health (Zhang and Chen 2020, p. 81).

5.1.4 Codes of Conduct (Self-Regulation)

In addition to legal regulation by the State and the creation of ethical standards by interested organizations, it is recommended that RS providers also act proactively in the implementation of systems that respect ethics and human rights. The creation of codes of conduct and ethical standards for the sending of recommendations by the platforms themselves may be an important self-regulation tool, also helping companies to comply with the law when it is effectively applied (Privacy International 2018, pp. 13–28).

An example in this regard was the creation of the "Partnership on Artificial Intelligence to Benefit People and Society", originally established by some of the big tech companies, such as Microsoft, Google, Amazon, Facebook and IBM, to study and formulate best practices for AI, in accordance with ethical principles (Privacy International 2018, p. 13). Among the objectives, it seeks to advance the public's understanding of technology, in addition to serving as a platform for discussion about AI and its possible impacts on people and society (Partnership on AI). However, it is crucial that these self-regulation codes and principles are effectively applied on practice.

5.1.5 Digital Education in AI

From citizens to top technology executives, society must be educated about the beneficial use, misuse and potential harm of AI, especially RS (European Parliament 2020, p. 84). It is critical that there is increased awareness of AI at all levels of education, in order to prepare citizens for the current digital age, making them better able to make informed decisions that will be increasingly impacted by technology (European Commission 2020c, p. 6).

In this context, the recent Digital Education Action Plan launched by the European Commission, to be applied between 2021–2027, is a good example of an educational project applicable to recommender systems. One of the main goals established was to improve the digital skills of citizens from childhood, which includes investing in basic knowledge of AI, ethical values associated with these technologies and awareness of the existence of digital rights (European Commission 2020a). Such measures would work as a relevant strategy for reducing information asymmetries, in addition to preventing risks by increasing public awareness, empowering users and the consequent effective exercise of rights.

The educational approach is even more important for private professionals who participate in the development processes of these technologies, as they must understand not only how to create accurate systems, but also build them in accordance with ethical and legal guidelines, based on human rights and democratic values. For example, another initiative encouraged by the European Commission is to transform some of the ethical principles into a "curriculum" to be followed by AI developers, as one of the stages of their training (European Commission 2020c, p. 6). Furthermore, whether through public or private initiatives, the development of ethics-related research in AI tools, such as RS, is essential.

5.2 Specific Legal Regulation for AI Systems

Due to the rapid implementation of RS and other AI's tools in different sectors, especially in digital platforms, and its harmful consequences, there are some initiatives to analyse possible forms of regulation of the technology, with especial

attention to the protection of vulnerable groups. To illustrate that, European Union's regulatory initiatives will be analysed as an example, given its potential to influence other regulations around the world due to the Brussels Effect, as occurred with the GDPR.

In this context, the regulation of disruptive technologies was first set through the establishment of ethical principles, guidelines and opinions on the development and use of AI, such as, for example, the 2019 Ethical Guidelines for Trustworthy AI by the Independent High-Level Expert Group on Artificial Intelligence – AI HLEG (2019) and the European Commission's White Paper on AI of February 2020. In this scenario, as mentioned in the previous topics, all stakeholders related to RS must pay attention to these ethical standards applicable to AI.

Yet, after the sedimentation of basic principles and guidelines applicable to AI, the EU is now trying to implement binding legal rules specifically applied to this technology, besides the already applicable data protection legislation, which the main example is the GDPR. Thus, recently, EU legislature approved and started the process of creation of legislations directed to AI and places where it is used (such as digital platforms). In the context of RS, the recent approved Digital Services Act (DSA) and, more directly, the proposal of Artificial Intelligence Act (AIA) are the most important examples.

5.2.1 Digital Services Act (DSA)

The DSA (2022) is an European Regulation that creates rules for the providers of certain information society services (digital services), especially through digital platforms. One of its innovative measures is the creation of rules that directly addresses recommender systems provided by online platforms. First, the regulation defines RS on Article 3 (s) as "*a fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service or prioritise that information (. . .)*", which is in line with the premise of Recital 70 that RS are the core part of the online platforms' business, since it facilitate and optimise access to information for the recipients of the service.

Consequently, as RS influences in the way the information flows in digital platforms,[6] the Regulation focus on the importance of transparency, creating on Recital 70 and Article 27 obligations related to the information required in digital platforms' terms and conditions (that should be written in plain and intelligible language) and options that these platforms must provide to the users in order to allow them to understand, modify or influence the recommendations' parameters. Also, specifically in the case of providers of very large online platforms and online search

[6] According to Recital 70, recommender systems of online platforms act algorithmically suggesting, ranking, prioritizing and curating information to facilitate the user's search of relevant content and improving user experience, besides the amplification of certain messages, the viral dissemination of information and the stimulation of online behaviour.

engines (article 33) that use RS—such as Meta and Google—article 38 require them to provide at least one option for each of their RS which is not based on profiling.

In that way, online platforms should consistently ensure that recipients of their service are appropriately informed about how recommender systems impact the way information is displayed and can influence how information is presented to them. They should clearly present the parameters for such RS in an easily comprehensible manner to ensure that the recipients of the service understand how information is prioritised for them. Those parameters should include at least the most significant criteria in determining the information suggested to the recipient of the service and the reasons for their respective importance.

As RS have a significant impact on people's behaviour and how they interact and find information online, the DSA intends to empower users through information and choice, enhancing GDPR's rules related to users' control over personal data. For example, the regulation sets obligation to providers of RSs of very large platforms to conduct risk assessments (article 34 (2) (a)), mitigate the risks founded through testing and adapting their algorithmic systems (article 35 (1) (d)) and explain, by the request of the European Commission or the Digital Service Coordinator, the design, the logic, the functioning and the testing of their systems (article 40 (3)).

Considering the problems related to RS, strengthening transparency obligations on online platforms and providing greater choice to users is an important first step to address the concerns fostered by this technology (Article 19 2021).

5.2.2 Proposal of an Artificial Intelligence Act (AIA)

The EU already has important regulation applicable to AI, such as GDPR, which provides some level of protection. However, according to the European Commission (2021), it was insufficient to address all the challenges that the technology may create, as saw in the previous topics. Thus, on April 2021,[7] the Commission proposed the first legal regulation specifically directed to AI, which aims to provide AI developers, deployers and user with clear requirements and obligations regarding the technology in order to both encourage innovation and protect potentially threatened fundamental rights and freedoms, creating an environment of trust.

The proposal is set in a risk-based approach, addressing the risks specifically created by AI applications, which may be considered unacceptable, high, limited or minimal to people's safety and fundamental rights. In accordance with Recital 14, although most AI systems existing today are considered of limited or minimal risk, being useful for society, depending on the intensity and the scope of the risks that AI may generate, it would be necessary to prohibit some AI practices; impose requirements for high-risk AI techniques and obligations for its operators; or also transparency obligations to certain AI systems.

[7] "Currently, the processing of the AI Act is in its final phase, following amendments by the Council of the European Union and the European Parliament"; Council of the European Union (2022).

Differently of what happens in the DSA, the AI Act Proposal does not specifically address recommendation systems, but it will inevitably apply to these tools, as they are based on AI and the generation of "recommendations" is covered by the Proposal's definition of AI on Article 3 (1) as one of its possible outputs.[8] Consequently, it is possible that recommendation systems will have a different treatment according to one of the four levels of risk they may create in the specific case.

With that said, at first, RS of minimal or no risk associated would be free to be developed and used. Yet, considering the potential manipulative uses, it may be prohibited when it is developed with "*subliminal techniques beyond a person's consciousness with the objective to or the effect of materially distorting a person's behaviour*"[9] or when it "*exploits any of the vulnerabilities of a specific group of persons due to their age, disability or a specific social or economic situation, with the objective to or the effect of materially distorting the behaviour of a person pertaining to that group*"[10] in a way that causes or is reasonable likely to cause physical or psychological harm.

In addition, there is a great chance that recommendation systems will be classified as high risk of harm to the health, safety or fundamental rights of individuals, according to the criteria of the AIA Proposal, defined on Article 6 and complemented by a list of high-risk application on Annex III.

If this is the case, high-risk recommender systems would be subject to a (third-party) conformity assessment with a series of obligations before they are put on the market or put into service—such as appropriate data governance (Article 10), elaboration of adequate risk management and mitigation systems (Article 9), technical documentation (Article 11), appropriate human oversight (Article 14) and provision of clear and adequate information to users (Transparency—Article 13)—but also would be subjected to enforcement after such RS is already in use. These ex-ante requirements related to transparency and risk-assessment would create an obligation to RS' providers to promote compliance by design in the case of high-risk recommender systems (Reinhold and Müller 2021).

Although the proposal has several memorable aspects, being the first regulation specifically directed to AI, serving as an international inspiration, there are still points of attention, such as the use of vague terms, the absence of an obligation to carry out a human rights impact assessment or the little mention of the possibility that people affected by AI systems have the power to challenge their harmful outcomes—with, for example, the establishment of the right not to be subject to

[8] Article 3 (1) of the Artificial Intelligence Act Proposal: "'artificial intelligence system' (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts".

[9] Article 5 (1) (a) of the Artificial Intelligence Act Proposal.

[10] Article 5 (1) (b) of the Artificial Intelligence Act Proposal.

a non-compliance AI system, right to explanation or the right to lodge a complaint with a supervisory authority) (Algorithm Watch 2022b).

For instance, if a RS has substantial effects on people's lives, it must not only be offered transparency concerning the implementation of the system, but mainly the possibility to challenge its decision (Reinhold and Müller 2021) (Algorithm Watch 2022a). Considering RS, thus, there must be legally and easily accessible options for affected people to question the recommendations and, if it is the case, to demand reversal, reconsideration through a different procedure, or even compensation.

In the case of RS of online platforms, through DSA, it is already possible to the users to modify or influence the main parameters of the system. However, mere technological solutions do not enough to ensure that AI systems are used in favour of the individuals, not just the providers. At this point, similar to what happened on DSA, accountability frameworks, empowering those directly affected by such systems, are an important aspect in this AI context (Reinhold and Müller 2021).

Furthermore, civil society still criticizes the last text of the AIA proposal, as there are yet some loopholes necessary for an adequate fundamental rights-based approach, especially in terms of meaningful accountability, public transparency and meaningful and balanced civil society participation (Algorithm Watch 2022a).

Thus, there is a current trend towards regulation of AI systems, such as recommendation systems, moving forward from a guidelines-principled approach in the direction of the development of binding legislative acts, as happens in the EU. However, it is necessary that these regulations do not act as a barrier to innovation, creating too rigid obligations, nor are they just the false appearance of regulation, creating vague and inoperative rules. Adequate regulation is essential for responsible innovation–which can be achieved with effective governance instruments, through regulation that is proportional to the systems' level of risk.

Recommender systems can fulfil a crucial role in democratic society and not only endanger, but also contribute to the realisation of fundamental rights and public values when well developed and used (Helberger et al. 2021). The new legislative initiatives must ensure that these systems work according to these values and not against it. Therefore, the union of the DSA and the proposed AIA may enhance users' empowerment and effective choice/control, mitigating potential risks and damages. It is a commendable first step, but we still have a long way to come.

6 Conclusion

In a hyperconnected world, with big data and information overload, recommender systems are increasingly present in our lives, silently predicting and inferring our interests, characteristics, and actions, influencing our decisions and categorizing us in behavioural profiles to send personalized content. Despite unquestionable benefits in terms of convenience, time management and organization, these tools pose considerable risks to fundamental rights, such as autonomy, privacy, data protection and non-discrimination.

Consequently, given the growing importance of these systems at the same time as the risk of adverse effects increases, there is a need for effective application and improvement of viable policies to face the multifaceted challenges they may cause. In other words, artificial intelligence applied to recommender systems must be regulated to prevent private interests from being privileged over the basic principle of "do not harm".

In this environment, GDPR represents a fundamental regulatory framework to address many of the human rights risks posed by the recommender systems' AI (Andersen 2018, pp. 30–31). As data is the engine of this technology, GDPR introduces a positive structure in favour of greater control of users over their data by establishing a series of rights, principles and requirements for the legal processing of personal data, especially in the case of automated decisions and creation of profiles. Many of these legal rules are drawn from ethical guidelines, based on human rights and values, such as transparency, justice, non-maleficence, beneficence, accountability, privacy, freedom, autonomy, dignity and solidarity, which are also fundamental to address the threats brought by RS.

These legal rules and ethical guidelines must also be reinforced by regulations coming from the technology itself, through "value-cantered design" strategies, where the architecture of RS considers these parameters in their way of functioning. Furthermore, for these tools to work in favour of the human being, it is also necessary to guarantee their adequacy based on impact assessments and algorithm audits, added to the establishment of codes of conduct by the market actors themselves. Besides that, "media literacy" policies are essential for the development of a society that will be able to understand the logic of these systems and, thus, make effectively informed decisions to reclaim control of their lives. Not least, the creation of specific regulation of AI systems or of their application environments, such as digital services provided by online platforms, is also essential to guarantee the good application of all these rules, since many of them will be integrated in these regulations.

Therefore, with the aim to maximize the benefits and mitigate the risks associated with RS, so that these tools are beneficial and not harmful to individuals and society, a multisectoral and multidisciplinary approach is essential, placing human being in the centre and involving all sectors of society, including contributions from ethical guidelines, technological functionalities, market self-regulation initiatives, educational policies and, to ensure effective application, the Law, Especially those directly created to the technology.[11]

[11] *See* generally, on the different applications of Machine Learning and, AI in this book A Oliveira and M A T Figueiredo—Artificial intelligence: historical context and state of the art; I Trancoso, N Mamede, B Martins, H S Pinto and R Ribeiro—The impact of language technologies in the legal domain; J Gonçalves-Sá and F L Pinheiro—Societal Implications of Recommendation Systems: A Technical Perspective; A T Freitas—Data-driven approaches in healthcare: challenges and emerging trends; M Correia and L Rodrigues—Security and Privacy; M Lanz and S Mijic—Risks associated with the use of natural language generation: Swiss civil liability law perspective; M S Fernandes and J R Goldim—Artificial Intelligence and Decision Making in Health: Risks

References

Ada Lovelace Institute, DataKind UK (2020) Examining the black box: tools for assessing algorithmic systems. Ada Lovelace Report, 29 Apr 2020. https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/. Accessed 15 Feb 2021

Aggarwal CC (2016) Recommender systems: the textbook. Springer International Publishing, Cham

Algorithm Watch (2022a) Civil society open letter demands to ensure fundamental rights protections in the Council position on the AI Act. https://algorithmwatch.org/en/fundamental-rights-protections-in-the-council-position-on-the-ai-act/. Accessed 27 Feb 2023

Algorithm Watch (2022b) A guide to the AI Act, the EU's upcoming AI rulebook you should watch out for. https://algorithmwatch.org/en/ai-act-explained/. Accessed 27 Feb 2023

Andersen L (2018) Human rights in the age of artificial intelligence. Access Now Report, Nov 2018. Accessed 15 Feb 2021

Article 19 (2021) EU: regulation of recommender systems in the Digital Services Act. Posted on 14th May 2021. https://www.article19.org/resources/eu-regulation-of-recommender-systems-in-the-digital-services-act/. Accessed 27 Feb 2023

Beil M, Proft I, Van Heerden D, Sviri S, Van Heerden PV (2019) Ethical considerations about artificial intelligence for prognostication in intensive care. Intensive Care Med Exp 7:70

Bernal P (2014) Internet privacy rights: rights to protect autonomy. Cambridge University Press, New York

Bigonha C (2018) Inteligência artificial em perspectiva. Panorama setorial da Internet. Intel Artif Ética 10:1–9

Bioni BR (2019) Proteção de dados pessoais: a função e os limites do consentimento. Forense, Rio de Janeiro

Bioni BR, Luciano M (2019) O princípio da precaução para a regulação da inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada? In: Frazão A, Mulholland C (eds) Inteligência artificial e direito: ética, regulação e responsabilidade. Editora Revista dos Tribunais, Sao Paulo, p 720

Bittencourt I, Gomes E (2019) O consentimento nas leis de proteção de dados pessoais: análise do regulamento geral sobre proteção de dados Europeu e da lei Brasileira 13.709/2018. In: Anjos L, Brandão L, Polido F (eds) Políticas, internet e sociedade. Instituto de Referência em Internet e Sociedade (IRIS), Belo Horizonte, pp 26–35

Bradford A (2012) The Brussels effect. Northwest Univ Law Rev 107(1):2012

Cavoukian A (2009) Privacy by design: the 7 foundational principles. Information and privacy commissioner of ontario. https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf. Accessed 20 Feb 2021

Council of the European Union (2022) Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence

and Opportunities; W Gravett—Judicial Decision-making in the Age of Artificial Intelligence; D Durães, P M Freitas and P Novais—The Relevance of Deepfakes in the Administration of Criminal Justice. *See* also, on Ethics, in this book P U Lima and A Paiva—Autonomous and Intelligent Robots: Social, Legal and Ethical Issues; A T Freitas—Data-driven approaches in healthcare: challenges and emerging trends; M C Patrão Neves and A B Almeida—Before and Beyond Artificial Intelligence: Opportunities and Challenges; M S Fernandes and J R Goldim—Artificial Intelligence and Decision-Making in Health: Risks and Opportunities; M N Duffourc and D S Giovanniello—The Autonomous AI Physician: Medical Ethics and Legal Liability; R Nogaroli and J L M Faleiros Júnior—Ethical challenges of artificial intelligence in medicine and the triple semantic dimensions of algorithmic opacity with its repercussions to patient consent and medical liability; and B A Ribeiro, H Coelho, A E Ferreira and J Branquinho—Metacognition, Accountability and Legal Personhood of AI.

Act) and amending certain Union legislative acts. Brussels, 25 November 2022. https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf

Currie G, Hawk KE, Rohren EM (2020) Ethical principles for the application of artificial intelligence (AI) in nuclear medicine. Eur J Nucl Med Mol Imaging 47:748–752

Edwards L, Veale M (2018) Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions?". IEEE Secur Priv 16:46–54

Ekstrand JD, Ekstrand MD (2016) First do no harm: considering and minimizing harm in recommender systems designed for engendering health. In: Engendering health workshop at the RecSys 2016 conference. ACM, Boston, pp 1–2

European Commission (2018) Communication from the commission to the European Parliament, the European Council, the Council, the European economic and social committee and the committee of the regions: Artificial intelligence for europe, COM(2018)237–communication. European Commission, Brussels

European Commission (2020a) Communication from the commission to the European Parliament, the European Council, the Council, the European economic and social committee and the committee of the regions. Digital education action plan 2021–2027: resetting education and training for the digital age, COM/2020/624 final. European Commission, Brussels

European Commission (2020b) Proposal for a legal act of the European Parliament and the Council laying down requirements for artificial intelligence (Ares(2020)3896535). European Commission, Brussels

European Commission (2020c) White paper: on artificial intelligence-a European approach to excellence and trust, COM(2020) 65 final. European Commission, Brussels

European Commission (2021) Regulatory framework proposal on Artificial Intelligence. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Parliament (2019) State of the art and future of artificial intelligence. Briefing requested by the IMCO committee. Policy Department for Economic, Scientific and Quality of Life Policies, Directorate-General for Internal Policies. European Parliament, Brussels

European Parliament (2020) The ethics of artificial intelligence: issues and initiatives. Panel for the Future of Science and Technology, European Parliament, Brussels

Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E (2018) AI4People-an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach 28:689–707

Guszcza J, Lee M, Ammanath B, Kuder D (2020) Human values in the loop: design principles for ethical AI. Deloitte Rev Technol Ethics 26:65–81

Hartmann IA, Franqueira BD, Iunes J, Abbas L, Curzi Y, Villa B, Abreu F, Dias R (2020) Regulação de inteligência artificial no Brasil: policy paper. Contribuição do Centro de Tecnologia e Sociedade (CTS) – Fundação Getulio Vargas (FGV Direito Rio) à Consulta Pública do Ministério da Ciência Tecnologia Inovações e Comunicações – MCTIC sobre a Estratégia Brasileira de Inteligência Artificial. FGV DIREITO RIO

Helberger N et al (2021) Regulation of news recommenders in the Digital Services Act: empowering David against the Very Large Online Goliath. Internet Policy Rev 26. https://policyreview.info/articles/news/regulation-news-recommenders-digital-services-act-empowering-david-against-very-large

High-Level Expert Group on Artificial Intelligence – AI HLEG (2019) Ethical guidelines for trustworthy AI. European Commission. https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf

Human Rights Watch (2018) The EU general data protection regulation: questions and answers. https://www.hrw.org/news/2018/06/06/eu-general-data-protection-regulation#. Accessed 10 Feb 2021

Jannach D, Zanker M, Felfernig A, Friedrich G (2010) Recommender systems - an introduction. Cambridge University Press, New York

Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nat Mach Intell 1:389–399

Kanoje S, Girase S, Mukhopadhyay D (2015) User profiling for recommender system. In: 4th Post graduate conference for information technology (iPGCon-2015). Amrutvahini College of Engineering, Sangamner

Magrani E (2014) Democracia conectada: a Internet como ferramenta de engajamento político-democrático. Jeruá – FGV Direito Rio, Curitiba

Magrani E (2019) Entre dados e robôs: ética e privacidade na era da hiperconectividade. Arquipélago Editorial, Porto Alegre

Magrani E, Silva P, Viola R (2019) Novas perspectivas sobre ética e responsabilidade de inteligência artificial. In: Frazão A, Mulholland C (eds) Inteligência artificial e direito: ética, regulação e responsabilidade. Revista dos Tribunais, São Paulo, p 720

Mazeh I, Shmueli E (2020) A personal data store approach for recommender systems: enhancing privacy without sacrificing accuracy. Expert Syst Appl 139:112858

Milano S, Taddeo M, Floridi L (2019) Recommender systems and their ethical challenges. AI Soc 35:957–967

Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L (2016) The ethics of algorithms: mapping the debate. Big Data Soc 3:2053951716679679

Monteiro RL (2018) Existe um direito à explicação na Lei Geral de Proteção de Dados do Brasil? Instituto Igarapé, Art. Estratégico 39. Adopted on Dez/2018

Mulholland C, Frajhof IZ (2019) Inteligência artificial e a lei geral de proteção de dados pessoais: breves anotações sobre o direito à explicação perante a tomada de decisões por meio de machine learning. In: Frazão A, Mulholland C (eds) Inteligência artificial e direito: ética, regulação e responsabilidade. Thomson Reuters Brasil, São Paulo, pp 267–292

Nadee W (2016) Modelling user profiles for recommender systems. Doctoral dissertation, Queensland University of Technology

Paraschakis D (2017) Towards an ethical recommendation framework. In: 11th International conference on research challenges in information science (RCIS). IEEE, Brighton, pp 211–220

Paraschakis D (2018). Algorithmic and ethical aspects of recommender systems in E-commerce. Doctoral dissertation, Malmö University

Pariser E (2011) The filter bubble: what the Internet is hiding from you. Penguin Press, New York

Privacy International (2018) Article 19: privacy and freedom of expression in the age of artificial intelligence. Apr 2018. https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf. Accessed 20 Feb 2021

Reinhold F, Müller A (2021) AlgorithmWatch's response to the European Commission's proposed regulation on Artificial Intelligence – A major step with major gaps. Algorithm Watch, published on 22 April 2021. https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021. Accessed 4 July 2021

Reisman D, Schultz J, Crawford K, Whittaker M (2018) Algorithmic impact assessments: a practical framework for public agency accountability. AI Now Institute. https://ainowinstitute.org/reports.html. Accessed 15 Feb 2021

Russell S, Dewey D, Tegmark M (2015) Research priorities for robust and beneficial artificial intelligence. AI Mag 36:105–114

Sahu S, Singh S (2019) Ethics in AI: collaborative filtering based approach to alleviate strong user biases and prejudices. In: 2019 Twelfth International conference on contemporary computing (IC3). IEEE, Noida, pp 1–6

Seaver N (2018) Captivating algorithms: recommender systems as traps. J Mater Cult 24:421–436

Silva PR (2020) Os direitos dos titulares de dados. In: Mulholland C (ed) A LGPD e o novo marco normativo no Brasil. Arquipélago Editorial, Porto Alegre, p 400

Souza CA, Perrone C, Magrani E (2021) O direito à explicação: entre a experiência europeia e a sua positivação na LGPD. In: Bioni B, Doneda D, Sarlet IW, Schertel L, Rodrigues OL (eds) Tratado de proteção de dados pessoais. Forense, Rio de Janeiro, pp 243–270

Sunstein CR (2007) Republic.com 2.0. Princeton University Press, New Jersey

Susser D, Roessler B, Nissenbaum H (2019) Technology, autonomy, and manipulation. Internet Policy Rev 8:22

Tejeda-Lorente Á, Bernabé-Moreno J, Herce-Zelaya J, Porcel C, Herrera-Viedma E (2018) Adapting recommender systems to the new data privacy regulations. In: Fujita H, Herrera-Viedma E (eds) New trends in intelligent software methodologies, tools and techniques. IOS Press, Amsterdam, pp 373–385

Van Ooijen I, Vrabec HU (2018) Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. J Consum Policy 42:91–107

Varshney LR (2020) Respect for human autonomy in recommender systems. arXiv preprint arXiv:2009.02603

Wachter S (2020) Affinity profiling and discrimination by association in online behavioral advertising. Berkeley Technol Law J 35:367

Wachter S, Mittelstadt B (2019) A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Colum Bus Law Rev 2019:494

WP29 (2014) Article 29 data protection working party. Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of directive 95/46/EC. Adopted on 9 Apr 2014

WP29 (2017) Article 29 data protection working party. Guidelines on automated individual decision-making and profiling for the purposes of regulation 2016/679. Adopted on 3 Oct 2017

WP29 (2018) Article 29 data protection working party. Guidelines on automated individual decision-making and profiling for the purposes of regulation 2016/679. Adopted on 6 Feb 2018

Zhang Y, Chen X (2020) Explainable recommendation: a survey and new perspectives. Found Trends Inf Retr 14:1–101

Zhang Q, Lu J, Jin Y (2020) Artificial intelligence in recommender systems. Complex Intell Syst 7:439–457