Indian Journal of Law and Technology

Volume 17 | Issue 1 Article 3

2021

# Recommender Systems and Autonomy: A Role for Regulation of Design, Rights, and Transparency

Christian Djeffal

Christina Hitrova



## Recommended Citation

Djeffal, Christian and Hitrova, Christina (2021) "Recommender Systems and Autonomy: A Role for Regulation of Design, Rights, and Transparency," *Indian Journal of Law and Technology*: Vol. 17: Iss. 1, Article 3.
DOI: 10.55496/JFSP9105
Available at: https://repository.nls.ac.in/ijlt/vol17/iss1/3

# Recommender Systems and Autonomy: A Role for Regulation of Design, Rights, and Transparency

*Christian Djeffal*[*], *Christina Hitrova*[**] *& Eduardo Magrani*[***]

**Abstract** *Recommender systems are now widely deployed across multiple dimensions of the digital reality that increasingly shapes our lives. In doing so, they mould individual thoughts and actions and can affect individual and collective autonomy. In this paper we first discuss how the ubiquitous exercise of 'soft' power by recommender systems on individual users presents interference into individual autonomy and its legal dimensions, expressed through collective and individual self-determination, democratic values and institutions, as well as individual human rights and freedoms. We then argue that this exercise of power over individual and collective destinies necessitates regulatory action to establish an appropriate system of checks and balances on recommender systems and their creators. Utilising a bottom-up approach, we look at the fundamental aspects of a recommender system's design and functioning that shape the impact these algorithms have on individual autonomy. On the basis of this, we identify three key areas where regulation can be targeted in order to empower users and address current power imbalances - (1) algorithmic design, (2) data protection rights, and (3) transparency and oversight. We map the key questions and options for future regulatory action in each of these domains, highlighting the decisions and competing interests that regulators will need to consider. We conclude by discussing the policy implications of this mapping of the debate and the relevance they have for the future of recommender systems regulation.*

[*] Christian Djeffal is Assistant Professor for Law, Science and Technology at the Technical University of Munich. He researches and lectures on the relationship between law and digital technology.

[**] Christina Hitrova works on Responsible AI with PricewaterhouseCoopers (Czech Republic) and previously researched and consulted stakeholders on the links between law, ethics and technology at the Technical University of Munich and The Alan Turing Institute.

[***] Eduardo Magrani is a Doctor of Laws and an Affiliate at the Berkman Klein Center for Internet & Society at Harvard University. He is also the President of the National Institute for Data Protection in Brazil and Partner at Demarest Advogados.

We would like to acknowledge the help of our research assistants Daan Herpers and Shivangi Mishra.

## I. Introduction

As the amount of information uploaded to the Internet has continued to grow, exploring content without any sort of structure or guidance has become overwhelming, possibly even impossible. Every second 6 new websites are published, 1,099 posts are shared on Instagram, 4,050 photos are uploaded to Facebook and 5,787 tweets are posted on Twitter. These numbers increase every second.[1] Mirroring this explosion, recommender systems ('RS') have quickly become ubiquitous and are currently used to personalise content choices and rankings across platforms and apps. RS are algorithms that curate — what they identify as — relevant information by tailoring it to individual users through data processing techniques. RS are used to recommend friends or content on social media, but they can just as easily

[1] Spectralplex, 'How Much Content is Uploaded to the Internet Per Second?' (*Spectralplex*) <https://spectralplex.com/how-much-content-is-uploaded-to-the-internet-per-second/> accessed 25 March 2021.

be used to suggest tailored diets or exercises in weight loss apps or present options for travel routes on the basis of traffic density information. Behind the scenes, RS are also used in targeted and behavioural advertising — the engine of dominant Internet business models. The profiling and user tracking needed for personalisation have been criticised due to the privacy intrusions that they give rise to. In this article, however, we argue that the impact of RS goes far beyond such privacy concerns. Instead, the suggestive power of 'recommendations' based on individual thoughts and actions can impact individual autonomy and, by extension, human rights as well as individual and collective self-determination.

Through their functioning, RS increasingly shape our experience in a virtual environment[2] and thanks to machine learning ('ML') and the increasing collection of personal data, these algorithms can now enable granular and persuasive micro targeting. This brings about tangible shifts in our thoughts and actions or, stated otherwise, autonomy. Individuals' choices could be affected by RS determining the content or options visible to them.[3] Individuals can also be influenced by the order in which information is presented;[4] we prioritise those items that are ranked higher on a list.[5] By doing this in the absence of conscious awareness and individual choice regarding how content is targeted at them, recommendations may disrupt an individual's capacity of self-determination. A recent, notorious example of the impact RS can have on individual autonomy, thoughts and actions, is that of Molly Russel. Molly was a fourteen-year-old schoolgirl who took her own life in November 2017, days before her fifteenth birthday. After her death, Ian Russel, her father, publicly blamed Big Tech, in particular Instagram,

---

2 Silvia Milano, Mariarosaria Taddeo and Luciano Floridi, 'Recommender Systems and their Ethical Challenges' (2020) 35 AI & Society 957.

3 Christine Clavien, 'Ethics of Nudges: A General Framework with a Focus on Shared Preference Justifications' (2018) 47 Journal of Moral Education 366.

4 Andreas Hellmann, Chiing Yeow and Lurion De Mello, 'The Influence of Textual Presentation Order and Graphical Presentation on the Judgements of Non-Professional Investors' (2017) 47 Accounting and Business Research 455; Buck KW Pei, Philip MJ Reckers and Robert W Wyndelts, 'The Influence of Information Presentation Order on Professional Tax Judgment' (1990) 11 Journal of Economic Psychology 119; Michael Eisenberg and Carol Barry, 'Order Effects: A Study of the Possible Influence of Presentation Order on User Judgments of Document Relevance' (1988) Journal of the American Society for Information Science 8.

5 Mark T Keane, Maeve O'Brien and Barry Smyth, 'Are People Biased in Their Use of Search Engines?' (2008) 51 Communications of the ACM 49; Jonah Berger, 'Does Presentation Order Impact Choice After Delay?' (2016) 8 Topics in Cognitive Science 670.

for his daughter's death.[6] After Molly's death, Mr. Russel found out that his daughter's Instagram newsfeed was full of suicidal posts. In his own words:[7]

> I think Molly probably found herself becoming depressed. She was always very self-sufficient and liked to find her own answers. I think she looked towards the internet to give her support and help. She may well have received support and help, but what she also found was a dark, bleak world of content that accelerated her towards more such content.

Mr. Russel alleged that Instagram's algorithms, by targeting content at Molly, ended up pushing her into that "dark rabbit hole of depressive suicidal content."[8] This case illustrates a broader phenomenon of social media influences and content curation that affect individual's physical and mental health,[9] and not just of children. Depending on their field of application, RS could also affect the rights of users, e.g., if used in news media or in healthcare. RS that prioritises some news or publications at the expense of others could be softly limiting the right of readers to access information or of writers to express their opinions and impart information. RS in health apps, by providing suggestions for exercise or diet, could directly play a role in the health of their users, thus affecting their right to health. The causal relationship between RS and particular outcomes for their users is soft but extant. Yet, assessing the power and influence RS have in a triangular relationship is trickier, e.g., situations where a RS shapes the information served to a user and it is the actions of that user that then go on to produce rights-impacting effects. For example, a RS used by a doctor could suggest a particular treatment for a patient, but it is the actions of the doctor that would ultimately determine what treatment is provided. Or, more controversially, a RS could present content against a particular protected demographic group (e.g., religious, racial, etc.) to people already demonstrating bigoted beliefs, and can, thus, encourage a view of the world that could potentially push them towards committing violence against members of those racial or religious

---

[6] Jacob Dirnhuber, 'Heartbroken Dad Claims Instagram 'helped to Kill His 14-Year-Old Daughter' Who took her Own Life after Viewing Suicide Posts' *The Sun* (22 January 2019) <https://www.thesun.co.uk/news/8258105/ian-russell-molly-instagram-killed-daughter/> accessed 31 March 2021.

[7] Press Association, 'Molly Russell Entered "Dark Rabbit Hole of Suicidal Content" Online, Says Father' *Evening Express* (17 January 2020) <https://www.eveningexpress.co.uk/news/molly-russell-entered-dark-rabbit-hole-of-suicidal-content-online-says-father-2/> accessed 31 March 2021.

[8] ibid.

[9] Faith Ridler, 'Now 30 Families Blame Social Media Firms for Their Roles in Children's Suicides' *Mail Online* (27 January 2019) <https://www.dailymail.co.uk/news/article-6636807/Now-30-families-blame-social-media-firms-roles-childrens-suicides.html> accessed 31 March 2021.

groups. Dissecting the causal role of RS in such triangular relationships is complex and most likely does not meet the legal standard of causation in most jurisdictions. Nevertheless, it is clear that RS play a significant role in shaping the perceptions, and thus scope for autonomy of their users.

The persuasive strength of the impact of RS on autonomy may range, at its most innocent, from small nudges to premeditated and targeted manipulation of information and individuals.[10] Nudging is the design of choice architecture that pushes individuals towards a predictable and desirable behaviour without explicitly limiting freedom of choice.[11] Instead, it does so by relying on "cognitive boundaries, biases, routines, and habits."[12] Even if not directly limiting choices, a choice architecture might interfere with the ability of a person to identify and consider their options and, thus, affects their agency.[13] The use of personal information can enhance the effectiveness of recommendations, increasing the 'controlling' power of influences on the individual, threatening autonomy.[14] In the digital realm, creating architectures that affect individual choices may be unavoidable; for example, there must be a choice of layout and user interface.[15] Some have highlighted that careful considerations are needed when acting as a 'digital choice architect,' due to the great impact such decisions have on user actions.[16] In a digital setting, people are more likely to act automatically or intuitively,[17] with decreased attention span and concentration,[18] while being increasingly distracted and

---

[10] Daniel Susser, Beate Roessler and Helen Nissenbaum, 'Technology, Autonomy, and Manipulation' (2019) 8 Internet Policy Review <https://policyreview.info/node/1410> accessed 5 March 2021.

[11] Richard H Thaler and Cass R Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Rev and expanded ed, Penguin Books 2009).

[12] Pelle Guldborg Hansen, 'The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove?' (2016) 7 European Journal of Risk Regulation 155.

[13] JS Blumenthal-Barby, 'Choice Architecture: A Mechanism for Improving Decisions While Preserving Liberty?' in Christian Coons and Michael Weber (eds), *Paternalism: Theory and Practice* (Cambridge University Press 2013).

[14] Susser, Roessler and Nissenbaum (n 10) 3.

[15] Daniel M Hausman and Brynn Welch, 'Debate: To Nudge or Not to Nudge*' (2010) 18 Journal of Political Philosophy 123, 124; Tobias Mirsch, Christiane Lehrer and Reinhard Jung, 'Digital Nudging: Altering User Behavior in Digital Environments', *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)* (2017) <https://wi2017.ch/images/wi2017-0370.pdf> accessed 5 March 2021; Cass R Sunstein and Richard H Thaler, '"Preferences, Paternalism, and Liberty"' (2006) 59 Royal Institute of Philosophy Supplement 233, 250.

[16] Tim-Benjamin Lembcke and others, 'To Nudge or Not to Nudge: Ethical Considerations of Digital Nudging Based on Its Behavioral Economics Roots' 18, 10.

[17] Shlomo Benartzi and Jonah Lehrer, *The Smarter Screen: Surprising Ways to Influence and Improve Online Behavior* (2015).

[18] Ziming Liu, 'Reading Behavior in the Digital Environment: Changes in Reading Behavior over the Past Ten Years' (2005) 61 Journal of Documentation 700.

multitasking.[19] Moreover, the digital environment offers a wealth of tools and options at the disposal of creators and designers, along with the ability to micro-target and personalise content which may increase the effectiveness of nudges.[20] Even when users are aware of the role that algorithms play in online settings, they remain confident about their own autonomy and do not account for how they might be influenced.[21] In fact, knowing that information, e.g. advertisement, is targeted specifically to us might even change the way we view ourselves and the qualities that we associate with ourselves.[22] All of this suggests that individuals may be more vulnerable to decision-making errors in the digital realm both due to traditionally studied biases, as well as due to digital-specific and visual biases.[23]

Despite their potential far-reaching impact, until now RS have operated with little regulation to ensure checks and balances on their influence. Their name – 'recommender systems' – leaves the impression that their impact on human lives is soft and superficial. However, their influence could be described as analogous to Nye's concept of soft power in international relations – "the ability to affect others to obtain the outcomes one wants through attraction rather than coercion or payment."[24] By shaping our attention, RS attract us to one action or another. In some circumstances, the effects of this attraction can be equated with a *de facto* force, as seen in the Molly Russel case.

However, this lack of regulatory framework is changing. There have been indications that regulating RS has been on the minds of policy-makers in Europe. In November 2021, the European Commission proposed an AI Regulation ('the (draft) AI Act')[25] that seeks to establish common *ex ante* market requirements and *ex post* control measures on AI systems to ensure their safety and trustworthiness. The AI Act includes software that generates

---

[19] Kep Kee Loh and Ryota Kanai, 'How Has the Internet Reshaped Human Cognition?' (2016) 22 The Neuroscientist 506.

[20] Lembcke and others (n 16) 8.

[21] Leyla Dogruel, Dominique Facciorusso and Birgit Stark, '"I'm Still the Master of the Machine." Internet Users' Awareness of Algorithmic Decision-Making and Their Perception of Its Effect on Their Autonomy' (2020) Information, Communication & Society 1.

[22] Christopher A Summers, Robert W Smith and Rebecca Walker Reczek, 'An Audience of One: Behaviorally Targeted Ads as Implied Social Labels' (2016) 43 Journal of Consumer Research 156.

[23] Lembcke and others (n 16) 8.

[24] Joseph S Nye, 'Public Diplomacy and Soft Power' (2008) 616 The ANNALS of the American Academy of Political and Social Science 94, 94.

[25] Commission, "Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts" COM (2021) 206 final.

influential recommendations in its definition of AI[26] and goes on to explicitly prohibit AI systems that use "subliminal techniques beyond a person's consciousness."[27] It also prohibits systems that exploit "vulnerabilities of a specific group of persons due to their age, physical or mental disability"[28] in order to "materially distort a person's behaviour" in a way that causes or is likely to result in physical or psychological harm to that person or others. This prohibition seems to echo the dangers demonstrated by the Molly Russell case. In December 2021, the EU Digital Services Act ('DSA') was proposed.[29] The DSA also pays special attention to RS, particularly as used by very large online platforms and in advertising, and provides for multiple pathways to enhance their transparency to end users, external auditors, and the general public.[30] The DSA also recognises the systemic risk that could arise from RS and imposes risk assessment and management obligations on very large platforms. The risks of disseminating illegal content, negative effects on the exercise of fundamental rights, including freedom of expression and information, and the automated misuse and manipulation of their services with the goal of affecting democratic processes and civic discourse were specifically highlighted.[31] Even though the DSA continues to develop, the European Parliament rapporteur and the Council have expressed a desire to further reinforce transparency and user control over RS, obligations on large platforms, search engines, and online market places.[32] In the same vein, in January 2022, the Cybersecurity Administration of China also published a set of regulations intended to regulate RS, pushing for greater user control, limits on what data the systems can use, as well as more transparency of how they function.[33]

What is clear from these recent legislative developments is that there is a movement towards tackling the challenges that RS have given rise

---

[26] AI Act, art 3(1).

[27] AI Act, art 5(1)(a).

[28] AI Act, art 5(1)(b).

[29] Commission, 'Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act)' and amending Directive 2000/31/EC, COM(2020) 825 final. (European Commission 2020) <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608117147218&uri=COM%3A 2020%3A825%3AFIN> accessed 30 March 2021.

[30] Digital Services Act (DSA), arts 29, 30.

[31] DSA, arts 26, 27.

[32] 'Legislative Train Schedule - Proposal for a Regulation of the European Parliament and the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC / After 2020-09' (*European Parliament*, 17 December 2021) <https://www.europarl.europa.eu/legislative-train> accessed 15 January 2022.

[33] Arendse Huld, 'China Passes Sweeping Recommendation Algorithm Regulations' (*China Briefing News*, 6 January 2022) <https://www.china-briefing.com/news/china-passes-sweeping-recommendation-algorithm-regulations/> accessed 15 January 2022.

to - whether it is manipulation, behavioural change, fundamental rights impacts, or whether it is a larger scale impact on democratic processes and civic discourse, which affect collective self-determination. As these legislative processes evolve, we will observe how interests intertwined in this topic are to be balanced against each other.

With this article we would like to contribute to this debate. We focus on the perspective of individual users of RS, excluding from our analysis triangular situations where RS support decisions made by users regarding other individuals, and explore, from a European legal perspective, what role regulation does and can play in empowering individual users and safeguarding individual autonomy. We direct our analysis to future regulatory directions along the axes of shaping user-centric RS design, enabling user control through a data protection rights-based approach, and facilitating informed decision-making and accountability through comprehensive transparency. Even though the challenges posed by RS are gaining attention around the world, this article is grounded in European legal developments and, therefore, future research would be needed to shed light on whether and how they might be suited for other legal systems.

In the following Section II, we introduce autonomy and self-determination and how they are manifested, implicitly or explicitly, in law. We explain how RS operate on a technical level and how this can affect key aspects of autonomy, both legally, as well as philosophically conceptualised. Then, in the rest of the article, we discuss the current state-of-the-art of safeguarding autonomy in regulatory frameworks and then highlight key options, decisions and pathways forward. The key areas for regulation that we discuss are algorithmic design, user data protection rights, and transparency and oversight of RS as they are either directly implicated in determining the way in which individual autonomy is affected, as in the case of algorithmic design, or they constitute valuable tools to empower users or their representatives to safeguard individual autonomy, as in the case of data protection rights and transparency.

Thus, in Section III.A, we discuss regulatory options for safeguarding and promoting autonomy in the design of RS, using a law-by-design approach. In Section III.B we focus on the privacy and data protection rights upon which individuals could rely in order to control the information about them used for profiling and recommendations. Finally, in Section III.C we discuss transparency as a vital tool to ease the current asymmetrical distribution of information and power between RS creators and users and as a key infrastructure to enable accountability and meaningful human oversight over the

power exerted by RS. The goal of our approach is not to limit or prohibit RS, as they may be a desirable feature of virtual environments. Instead, we seek to highlight the gaps and needs that a regulatory framework should seek to fill in order to foster the creation and use of RS in a manner that truly ensures individual autonomy – that users are in the driver's seat, that they are aware of and can themselves shape or even exclude recommendations from their virtual worlds as they see fit, and that, vitally, there is a clear legal recognition of the impact that RS have on individuals and society that demands responsibility.

## II. Autonomy and recommender systems

Autonomy is a normative concept about the rightful claim to self-determination in multiple contexts, be they collective or individual, national or international, and is also the foundation of human rights, democracy and the rule of law. The concept of autonomy in the law refers to the scope and ability of individuals or groups to make decisions for themselves or to "follow their own life plan"[34] and this right is explicitly or implicitly protected at multiple levels in the law. On a fundamental level, autonomy is demonstrated through the capacity of individuals to freely bind themselves in contracts, a manifestation of their self-determination.[35] On a higher level, autonomy is protected and enabled through legal certainty and the rule of law. Compliance with the rule of law makes governmental actions and the legal framework predictable and empowers individuals to plan their lives around them. This relation is enabled by transparency allowing individuals to judge the legality of their actions or claims. Autonomy is also a cross-cutting and transversal principle that is ingrained in every human right but comes to the fore in specific constellations, as demonstrated in human rights law practice. If a specific right is protected, this necessarily includes the autonomy of humans to use that right freely. The right to property includes the autonomy of a subject to dispose of property in any way, including destroying it. The freedom of opinion grants the right to make one's opinion known or to stay silent. Other rights are more clearly linked to individual autonomy, for example as art 8 of the European Convention on Human Rights (ECHR), the right to respect for private and family life. The European Court of Human Rights (ECtHR) interpreted this provision to find a right to personal autonomy, identity and integrity within

[34] Emily Jackson, *Regulating Reproduction: Law, Technology, and Autonomy* (Hart Pub 2001) 2.

[35] Thomas Gutmann, 'Some Preliminary Remarks on a Liberal Theory of Contract' (2013) 76 Law and Contemporary Problems 39.

art 8.[36] Human rights bills protecting human dignity also situate autonomy in this context. The right to self-determination, as enshrined in arts 1 of the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR), also shows that autonomy has a collective side. This side is expanded on especially in the area of minority rights, where a degree of autonomy could allow minority groups sufficient self-determination to benefit from the group rights that majorities typically experience.[37] The collective self-determination of peoples, communities, and nations is also linked to democratic institutions and processes. This points back to the literal historical meaning of autonomy. It means that individuals or groups should provide for the rules governing them. In contrast, the concept of heteronomy provides that the rules are made by somebody else.[38] Thus, individual and collective autonomy permeates and acts as a foundation to multiple layers of the legal order. This demonstrates that autonomy is a fundamental value and backbone of many legal systems that is safeguarded through the rule of law, fundamental rights, democratic processes and institutions, and even individual responsibility, liability, and the freedom to contract. As a cornerstone of many legal structures, there is an acknowledgement that autonomy is valuable and should be appropriately safeguarded. We will now explore whether and how RS interact with autonomy before exploring what the current and future legal landscape of regulating this relationship looks like in the next section.

In order to demonstrate how precisely RS and autonomy are interlinked, we take a conceptual approach towards understanding autonomy. Autonomy, in its practical ethical dimension, can be seen to require two essential conditions: independence from controlling influences (liberty) and capacity to intentionally act and decide (agency).[39] RS can affect both of these dimensions of autonomy on an individual and collective level. Relevant to RS, exerting control and influence or manipulation can affect the decision-making capacity of individuals,[40] thus making them subject to the will

---

[36] Jill Marshall, *Personal Freedom through Human Rights Law? Autonomy, Identity and Integrity under the European Convention on Human Rights* (Martinus Nijhoff Publishers 2009) .

[37] J Wright, 'Minority Groups, Autonomy, and Self-Determination' (1999) 19 Oxford Journal of Legal Studies 605.

[38] Simon Hornblower, 'Autonomy' in Tim Whitmarsh (ed), *Oxford Classical Dictionary* (Oxford University Press 2015)..

[39] Lav R Varshney, 'Respect for Human Autonomy in Recommender Systems' [2020] arXiv:2009.02603 [cs] <http://arxiv.org/abs/2009.02603> accessed 5 March 2021; Tom L Beauchamp and James F Childress, *Principles of Biomedical Ethics* (8th edn, Oxford University Press 2019).

[40] Beauchamp and Childress (n 39).

of another.[41] This could affect individual liberty directly, by limiting the scope for individual decision-making, or more perniciously, by distorting the individual capacity to make informed decisions and, thus, their agency. For example, RS have been linked to the creation of 'filter bubbles' that limit the range and diversity of information users see[42] and can lead to political polarisation and a partial view of the world. In an extreme form, filter bubbles could reinforce messages of suicide,[43] radicalization and extremisation,[44] and mistrust of vaccines,[45] thus affecting both individuals and whole communities. Such actions can also lead to direct consequences on human rights. The Molly Russel incident spoke to children's rights, especially the physical integrity of children. Healthcare treatment or diagnosis recommendations can touch upon these and the right to health. When used in the context of social media and content curation or moderation, RS can have an impact on democracy, the freedom of speech and personality rights. This shows that RS can affect a multiplicity of human rights and freedoms and shape the space within which individuals can act with true autonomy just by directing individual attention and shaping individual thoughts and actions. From an individual level, through humans as an intermediary, RS could bring about tangible effects on human rights, freedoms, but also democratic processes and collective self-determination. Despite these effects, the current legal framework does not reflect satisfactory safeguards of such interferences.

RS are specifically created for the purpose of shaping human behaviour by exerting soft but persistent influences on individual liberty and shaping the information available for exercising agency and independent decision-making. Explicit safeguards against the influence of RS over individuals may be

---

[41] Andreas T Schmidt, 'The Power to Nudge' (2017) 111 American Political Science Review 404.

[42] Tien T Nguyen and others, 'Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity', *Proceedings of the 23rd International Conference on World Wide Web - WWW '14* (ACM Press 2014) <http://dl.acm.org/citation.cfm?doid=2566486.2568012> accessed 5 March 2021; Engin Bozdag, 'Bias in Algorithmic Filtering and Personalization' (2013) 15 Ethics and Information Technology 209.

[43] David D Luxton, Jennifer D June and Jonathan M Fairall, 'Social Media and Suicide: A Public Health Perspective' (2012) 102 American Journal of Public Health S195.

[44] Philip Baugut and Katharina Neumann, 'Online News Media and Propaganda Influence on Radicalized Individuals: Findings from Interviews with Islamist Prisoners and Former Islamists' (2020) 22 New Media & Society 1437; Mark Alfano and others, 'Technologically Scaffolded Atypical Cognition: The Case of YouTube's Recommender System' (2020) Synthese <http://link.springer.com/10.1007/s11229-020-02724-x> accessed 5 March 2021.

[45] Deena Abul-Fottouh, Melodie Yunju Song and Anatoliy Gruzd, 'Examining Algorithmic Biases in YouTube's Recommendations of Vaccine Videos' (2020) 140 International Journal of Medical Informatics 104, 175; Harald Holone, 'The Filter Bubble and its Effect on Online Personal Health Information' (2016) 57 Croatian Medical Journal 298.

necessary because there is an inherent misalignment of interests between those designing and deploying RS and the final users. Creators of RS may seek to further their own commercial goals under a veneer of providing better service and more relevant content for users. Recent fieldwork with US developers of RS identified a common goal of 'hooking' people and keeping them on a particular platform,[46] which was reflected in the way that the RS was created. RS design and operation can take a multitude of shapes; designer intent and desire can play a significant role in how RS ultimately operate. While RS can be valuable to online users, they need to balance the interests of designers and users to be a valuable solution.[47] Due to the asymmetry of power and knowledge between designers and users in shaping and understanding RS, a balance may be difficult and unlikely without some form of regulation. Here, we briefly introduce the main architectures used in RS, the data needed, as well as the role of a desired target variable for which RS optimise. All of these features can have implications for the impact the system has on individual autonomy. Furthermore, these features are also currently within the exclusive domain of determination of RS designers and developers.

In order to operate, RS require definitions of what the range of options they can recommend are, what a 'good' recommendation is and how to identify it (i.e. a target variable which the systems seek to maximise), and how their performance is evaluated,[48] which allows for future improvements of RS. There are a number of commonly used RS techniques that allow (semi-) automation of recommendations. First, *collaborative filtering* focuses on how multiple users have historically rated items, in order to predict ratings of these items by other users who have not yet rated them. RS can do that by grouping either users or items together, on the basis of similarity metrics. The RS can then suggest content on the basis of what similar users liked or on the basis of what items are similar to what a user and other similar users have liked in the past.[49] In contrast, a *content-based RS* models a user's interests by analysing attributes of items that a specific user has interacted with in the past, focusing on the user's own behaviour to predict the user's future rating of a new item.[50] Finally, the *knowledge-based approach* invites users to directly specify their interests or requirements. These interests are then

[46] Nick Seaver, 'Captivating Algorithms: Recommender Systems as Traps' (2019) 24 Journal of Material Culture 421.
[47] Francesco Ricci, Lior Rokach and Bracha Shapira (eds), *Recommender Systems Handbook* (Springer US 2015) 6.
[48] Milano, Taddeo and Floridi (n 2) 2.
[49] Charu C Aggarwal, *Recommender Systems* (Springer International Publishing 2016) 8.
[50] ibid 14.

combined with the system's pre-programmed domain knowledge to generate recommendations.[51] An example could be exploring real estate websites that allow refining search results through numerous user-chosen filters. In reality, RS often use hybrid architectural approaches.

RS also need different types of data about users and content or items to operate, depending on the recommender technique used.[52] Data can be used *inter alia* to assess the user's interest in an item or to assess the similarity of different users or of different items. RS can rely on both explicit and implicit user feedback. Actions, e.g. recording users 'liking' or 'sharing' a piece of content or even visiting a page, can serve as an implicit positive rating of that content that RS then use to inform and reinforce their operation. More complex models can also include data about time and duration of interactions of users, location, social or network information, as well as external domain knowledge.[53] Demographic data classifiers can especially boost the accuracy of other RS techniques.[54] Clearly, the three RS architectural approaches, as well as the data used ascribe a different weight to a single user's actions in terms of their impact on determining the ultimate recommendations that user receives.

A third key feature of RS is their determination of what 'good' recommendations are. "Good" is an inherently subjective term, especially in the context of personalisation. What is a "good" recommendation for one would not be so for another. Moreover, to automate the computation and presentation of recommendations, 'good' needs to be defined mathematically. RS are said to present items that are of interest or relevant to a particular user.[55] but how that should be translated into the RS's design and what they should optimise for is not predetermined. In machine learning (ML), the technology behind many RS, this is the key role of a target variable – a specific and measurable variable that allows the ML model to calculate and predict whether its performance (recommendation) will be poor or good, based on data from past performance. In RS, the target variable is the measurable variable that designers have determined to be a good proxy measure of a "good" recommendation – e.g., whether a user interacts with a piece of content, whether they share it, whether they 'like' it etc. The RS then seeks to maximise this. The difficulty here lies in identifying which measurable variable(s) can be used to represent a user's positive reaction to a recommendation.

---

[51] ibid 15.
[52] Ricci, Rokach and Shapira (n 47) 9.
[53] Aggarwal (n 49) 2.
[54] ibid 19.
[55] See Ricci, Rokach and Shapira (n 47) 1.

The goals of service providers could be to increase the number or diversity of items sold, interacted with, increase user satisfaction and fidelity, or simply improve understanding of what the user wants.[56] The use of such a variable will not always coincide with the users' definition of what a 'good' recommendation is for them. Moreover, should a "good" recommendation only be assessed on the basis of how a user perceives it or is there also space for reflecting the reputability of a source or the content of the item being recommended? Imagine the case that a person positively reacts to a piece of content, advocating for racial inequality. Does that mean this was a "good" recommendation? The interests of designers and the users seem to be satisfied with this recommendation being rated positively, but is there space to discuss collective self-determination and the social interest? These are not easy questions and there is not necessarily one correct answer. But they are, nevertheless, decided every time a RS is created. For this reason, in the next section we start our mapping of the pathways to regulating RS in order to safeguard individual autonomy by first discussing the importance of regulating RS design and the role users can and should play within it.

## III. Regulating autonomy in recommender systems

Given the deep sources of tension between RS and autonomy, in this paper we seek to explore the potential for regulatory interventions to safeguard autonomy by enhancing user empowerment in three ways: (i) through the design and functioning of RS, utilising a law-by-design approach; (ii) through privacy and data protection rights to control RS data inputs, with a rights-based approach; and (iii) through transparency in user interactions and co-shaping of RS, with a process-based solution. We structure our analysis along these three dimensions— algorithmic design, data inputs, and transparency — to introduce specific regulatory options. At every step, we explore how autonomy is safeguarded in law today and the current state of art and propose pathways for the future, as possible solutions to further enhance autonomy.

### A. Designing Recommender Systems

New technologies of ML have fuelled the capacity of RS to increase their performance and issue more fitting recommendations. Generally, the improvements are due to the availability of training data allowing the respective algorithms to be optimized in certain regards. However, the design goals and the respective metrics towards which such an algorithm can be optimised

[56] ibid 5–6.

vary. They range from engagement with the RS by spending time or buying products to more general goals like accurate user information or presentation of information the user might not have been exposed to. Every system communicates to a person in order to support and shape their decision-making. Considering that RS exert such an important influence on persons, the argument can be made that there should be some ways for users to actively influence them or, at the very least, there should be some expectation on the part of the creators of RS to consider and account for the impact their work might have on the autonomy of the intended users. Here we will explore how regulation can influence algorithmic design and propose options to shape RS design in an autonomy-enhancing manner through law.

### i. State of the art of design for autonomy

In law, there is a growing amount of legislation that directly engages in the design process. Notable and known examples relate to privacy and IT-security, as provided for in arts 25 and 32 of the General Data Protection Regulation (GDPR), which lay down privacy and security by design obligations.[57] Thereby, they transfer legal principles into the very design of technologies by mandating they be considered at every step of a technology's creation, use, and maintenance. Data protection and IT-security are to be included in design processes as design goals of their own right, although ones of many, balanced against qualifications like the cost of implementation. This begs the question regarding whether regulators could add autonomy as another design goal in the same fashion as data protection and privacy. An analysis of the law shows that there are already first signs of including such design goals.

Take for example art 29 of the draft EU DSA. This provision specifically addresses RS in the context of online platforms. The transparency obligation in this article hints at a nascent autonomy by design principle. It provides that "*[v]ery large online platforms that use recommender systems shall set out in their terms and conditions …. any options for the recipients of the service to modify or influence those main parameters that they may have made available, including at least one option which is not based on profiling…*". The obligation states that users are empowered to modify the

[57] Peter Schaar, 'Privacy by Design' (2010) 3 Identity in the Information Society 267; Dag Wiese Schartum, 'Making Privacy by Design Operative' (2016) 24 International Journal of Law and Information Technology 151; Privacy and security by design obligations are also found in the current draft Indian data protection legislation. Saumyaa Naidu and others, 'The PDP Bill 2019 Through the Lens of Privacy by Design' (The Center for Internet & Society 2020) <https://cis-india.org/internet-governance/blog/the-pdp-bill-2019-through-the-lens-of-privacy-by-design>.

parameters of a RS. It indicates that, at the least, users can choose to receive recommendations not based on personal profiling. This could mitigate the role of personalisation that otherwise enhances the effectiveness of recommendations in achieving their pre-determined goal. Moreover, the DSA also seeks to establish a duty on the part of large online platforms to manage systemic risks arising from their platforms. The draft act mandates that large online platforms should particularly take into account the negative effects on fundamental rights, including the right to privacy, freedom of expression and information, non-discrimination, and the rights of the child[58] among these risks. Platforms are explicitly tasked to consider 'how their recommender systems and systems for selecting and displaying advertisement influence any of the systemic risks'.[59] They are then tasked with taking appropriate action to mitigate identified risks, including by altering how their RS operate,[60] and their risk management activities are subject to independent audits[61] and public disclosure.[62] This provision will affect not only the process of RS design, but also its long-term maintenance and review. It is geared towards pushing platforms to reflect on and mitigate risks that arise as a result of their functioning and, especially, of the design and operation of their RS and advertisement systems. Both of these requirements in the DSA clearly indicate that legislators are looking into checking the power and influence of RS, including by demanding that user autonomy is considered and enhanced by design.

Similarly, the recent European Commission proposal for an AI Act, while it does not directly address RS, or user autonomy, demonstrates that it seeks to safeguard individual autonomy in the face of powerful artificial intelligence (AI). Art 5(1)(a) of the AI Act seeks to ban any "system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behavior", while art 5(1)(b) addresses systems that exploit "any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behavior of a person pertaining to that group".[63] These prohibitions of systematic utilisation of weaknesses of individuals clearly address limitations in their capacity to exercise autonomy and highlight awareness of the persuasive powers of AI.

---

[58] DSA, art 26(1)(b).
[59] DSA, art 26(2).
[60] DSA, arts 27(a) and 27(b).
[61] DSA, art 28.
[62] DSA, art 33.
[63] Both alternatives are only applicable when applied "in a manner that causes or is likely to cause that person or another person physical or psychological harm" – arts 5(1)(a) and 5(1)(b).

At the national level, an explicit example of inclusion of autonomy by design can be found in the German Digital Healthcare Act.[64] This act supports digital technologies like mobile apps by providing for funding schemes from health insurances. The Digital Healthcare Act introduces Section 20(k)(1) of the German Social Law Book V,[65] which provides for measures to enhance patients' self-determination when it comes to digital applications and telemedicine. Section 139(e)(2) of the German Social Law Book V provides for requirements for health insurances to remunerate digital applications if the applications meet a set of criteria. One of these criteria are positive effects on healthcare. The respective draft secondary legislation mentions "patients' sovereignty" as one of the decisive criteria of positive effects. Thereby, patients' sovereignty is one of the evaluation criteria that designers of those apps would have to take into account even at the design stage if they want their app to be covered by health insurance.

An autonomy-by-design requirement could take different forms. It could require RS designers to mitigate and minimise the risks their systems pose to autonomy or it could require them to consider how to maximise and proactively help realise individual autonomy in the design of their technologies.[66] One example for the latter approach would be the technology clause in art 4(g) of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) which obliges states "*[t]o undertake or promote research and development of, and to promote the availability and use of new technologies, including information and communications technologies, mobility aids, devices and assistive technologies, suitable for persons with disabilities, giving priority to technologies at an affordable cost*".[67] This clause explicitly addresses the progressive realisation of autonomy in technology and shows that in very specific cases the law can demand or incentivise autonomy in technology. There are clearly instances of both logics in existing legislation at multiple levels. Regulators should determine which approach would be best-suited for their goals, perhaps taking a diversified view depending on the application of RS.

---

[64] DVG 2019 (BGBI I p 2562)

[65] SGB V 1988 (BGBI I p 2477).

[66] Wolfgang Hoffmann-Riem, 'Re:Claim Autonomy, Die Macht Digitaler Konzerne' in Jakob Augstein (ed), *Reclaim Autonomy: Selbstermächtigung in Der Digitalen Weltordnung* (Erste Auflage, Originalausgabe, Suhrkamp 2017) 122.

[67] Convention on the Rights of Persons with Disabilities (adopted 24 January 2007 UNGA A Res 61/106 (CPRD)), art 4(g); An example in the Indian context could be the digital accessibility provisions under the Rights of Persons with Disabilities Act 2016, for reference see, 'Digital Accessibility in the Rights of Persons with Disabilities Act 2016' (2017) Centre for Internet and Society, India.

Regardless, these examples signify the general trend in legislation to consider autonomy, sovereignty and self-determination of users in the context of algorithmic design. However, these examples are – to date – rather general. Therefore, the question arises—what regulatory possibilities are there to apply autonomy-by-design principles in real life? The regulation of RS to enhance autonomy can include a number of considerations and principles for designers to keep in mind, however, a case-by-case approach would be necessary to assess how precisely such principles are to be transposed into algorithmic design. This is due to the fact that the actual risks of RS can vary considerably, depending on their context of application. What tools might exist to help fulfil this? What regulatory structures might be relevant to establish in order to facilitate this? In the remainder of this section, we look at more concrete opportunities for this.

### ii. Ways to further enhance autonomy

Autonomy can be included as a regulatorily-mandated design goal for RS, as discussed. This would require a clarity of whether its goal is to minimise negative impact on autonomy or also to maximise the positive and empowering impact on individual autonomy. Beyond this, however, regulatory options can distinguish between setting autonomy as a design goal to be implemented throughout the process of technology creation, or rather focusing on the final impact of the technology on autonomy, perhaps through requiring that it meet desirable standards for access to the market. In order to have a better grasp on these choices, it is necessary to know about different concepts in the design of RS, as well as the links between algorithmic design and individual autonomy which will be explained below.

#### *a. User capacity and shared decision-making*

A fundamental starting point for autonomy-enhancing algorithmic design is a greater understanding of the factors that make up an autonomous human decision in a human-machine interaction. Interdisciplinary research is necessary to understand the conditions under which a human decision could be assumed to be independent. This is particularly important in order to delimit whether a certain system is considered a recommender system or whether human autonomy has shrunk so far that the system effectively operates as an automated decision-making system. Several criteria have been introduced as a delimitation. The competence of human recipients of recommendations is one of them.[68] Another question is the extent to which a decision could

[68] Philip Scholz, '22' in Spiros Simitis, Gerrit Hornung and Indra Spiecker Döhmann (eds), *Datenschutzrecht: DSGVO mit BDSG* (Nomos 2019) para 27; Mario Martini, 'Art. 22'

actually be influenced.[69] So far, understanding the factors affecting human capacity has been especially relevant in the context of current regulatory approaches to automated decision-making systems processing personal data, specifically as per Article 22 GDPR that lays down safeguards for such systems. Debates about when a system fulfils the definition of an automated decision-making system have pushed such discussions forward. The answer is necessarily binary: either a system is an automated decision-making system in the sense of art 22 GDPR or the provision does not apply. This is relevant because, when discussing how different design approaches could be used to enhance or mitigate impacts on individual autonomy, a greater understanding of the relationship between suggestions and recommendations and individual decision-making is necessary. For example, in the case of decision support systems, one could design a taxonomy that describes different levels of human-computer interaction ranging from simple filters to an automated decision-making system. One factor guiding the different levels of such a taxonomy could be the degree of autonomy that rests with the user when interacting with the system, which could help assess the risk of direct interaction with RS. Potential measures could be linked to the different level. Such a taxonomy could describe the different levels of human autonomy in the same way that levels of autonomy of humans are described for automated vehicles. At the very least, such understanding would be necessary to ground all subsequent regulatory and design activities around autonomy-by-design for algorithms.

### *b. Serendipity and randomization*

A technical aspect that directly shapes how and with what aim a recommendation nudges individuals is the process of choosing a target variable and optimising RS. The optimisation process is crucial in machine learning.[70] Setting the goals of optimisation is a key component of designing algorithms and an instance in which human agency can guide the way in which

in Boris Paal and Daniel Pauly (eds), *Datenschutz Grundverordnung: DS-GVO* (2nd edn, CH Beck 2018) para 18; Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (2018) WP251rev.01 29 <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053> accessed 31 March 2021.

[69] Gerald Spindler and Anna Z Horvátg, 'Art. 22 Automatisierte Entscheidungen Im Einzelfall Einschließlich Profiling' in Gerald Spindler and Fabian Schuster (eds), *Recht Der Elektronischen Medien* (CH Beck 2019); Article 29 Data Protection Working Party (n 68) 21.

[70] Suvrit Sra, Sebastian Nowozin and Stephen J Wright (eds), *Optimization for Machine Learning* (MIT Press 2012).

AI-systems operate.[71] In supervised learning, designers also specify a concrete and measurable target variable that the algorithm is trained to seek to optimise for.[72] While some tasks are binary like image recognition of a horse (it is or it isn't a horse), RS have to carry out more complex computations to predict whether showing a particular item to a user and at a particular order will result in a higher or lower target variable. Above, we gave the example of optimising for selling more or diverse items, but a target variable need not perpetuate the commercial interests of the designers. A target variable could also, for example, be used to mitigate the intentionally nudging impact RS can have on user autonomy by introducing unexpected recommendations in different ways – through serendipity, diversity, or randomisation.

The concept of serendipity centres around the question of how to recommend information that fits the interests of the respective person without recommendations being known or expected.[73] Through item-based grouping, a RS could help individuals with very obvious choices. A person looking for a hammer will probably need nails. However, a more complex RS might also be able to suggest a new system to hang something without damaging the wall. A similar but distinct concept is diversity. Unlike serendipity, diverse recommendations are not aimed at finding what the user is looking for in the first place. Rather, they confront the user with content that is different from what she or he expected.[74] A system recommending job ads might include ads that go beyond the imagination of the user, but which might also fit.

Finally, in contrast to serendipity, randomisation does not relate to the actual fit of a recommendation to a user but selects alternatives outside of what is recommended by the system. Randomisation can enhance the scope of action of a system by allowing it to confront people with data outside of the usual training. If a news RS is personalised in a way that operates as a filter bubble, curating content along a specific political stream, randomization might break that up by including recommendations beyond the confines of what the system can predict will be positively received by the user. In addition to enhancing the independence of users from the 'will' of the designers of RS, randomisation can be a valuable and desired feature for risk management systems and applications of RS in public bodies exercising some form

---

[71] Björn Haferkamp, 'Was Ist Optimal? Nutzen Und Fallstricke Der Optimie' in Björn Bergh (ed), *Big Data und E-Health* (Erich Schmidt Verlag 2017).

[72] David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51 University of California, Davis 653.

[73] Aggarwal (n 49) 3–4.

[74] Natali Helberger, Kari Karppinen and Lucia D'Acunto, 'Exposure Diversity as a Design Principle for Recommender Systems' (2018) 21 Information, Communication & Society 191.

of oversight. One example is the system used by the German tax authority to identify tax applications and recommend them for further human scrutiny. Section 88 of the German Tax Code provides for the necessity of a randomized human control of this recommender system. One measure to complement the automated risk assessment is the random selection of cases for human review,[75] irrespective of their risk level. This measure fulfils two functions.[76] First, it tests overall compliance, especially of the applications with low risk levels. Second, it allows for the evaluation of the system itself, as the risk management system should select certain applications in a random fashion for further review irrespective of their risk level. An additional step in the system used by the German tax authority is the freedom of users to completely sidestep it. Another requirement of the system is that officials must have complete access to all applications and must be able to select cases themselves. Thus, there are technical features which could serve as 'breaks' along the way from designer intent to user nudging, thus limiting the intentional influence RS exert on their users.

### *c. User control*

While the steps in the previous section demonstrate ways in which the link between designer interests and user influence through RS can be limited, user autonomy and self-determination can also be enhanced through greater participation of users in the shaping of the RS they use – user control. There is a vivid area of research that looks into whether and how users can influence RS voluntarily. Currently, individuals can and often do contribute to their information curation, e.g. by choosing whether to follow certain individuals, pages, channels or by blocking content from sources.[77] However, user control approaches go far beyond the ordinary acts of users providing profile data or giving feedback.[78] Instead, user controls entail more direct impact, e.g. settings through which users actively tweak and change the underlying

---

[75] Ann Cavoukian, 'The 7 Foundational Principles' (2009) <https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf> accessed 31 March 2021.

[76] Nadja Braun Binder, 'Ausschließlich Automationsgestützt Erlassene Steuerbescheide Und Bekanntgabe Durch Bereitstellung Zum Datenabruf' (2016) Deutsche Steuer-Zeitung 526.

[77] Lisa Merten, 'Block, Hide or Follow—Personal News Curation Practices on Social Media' (2020) Digital Journalism 1.

[78] For examples of user control see Yucheng Jin, Bruno Cardoso and Katrien Verbert, 'How Do Different Levels of User Control Affect Cognitive Load and Acceptance of Recommendations?' 2017 <http://ceur-ws.org/Vol-1884/paper7.pdf> 8; Mechanisms of instant feedback are described in Harald Steck, Roelof van Zwol and Chris Johnson, 'Interactive Recommender Systems: Tutorial', *Proceedings of the 9th ACM Conference on Recommender Systems* (ACM 2015) <https://dl.acm.org/doi/10.1145/2792838.2792840> accessed 5 March 2021.

algorithms[79] or can choose between different algorithms.[80] It puts users in the driver's seat and enhances their autonomy. So far, the reported results of experiments are promising. Users make active use of these possibilities, they have a positive experience[81] and such measures generally also increase their trust.[82] Therefore, user control is a design choice that can substantially add to recommender systems enhancing autonomy. As discussed above, the draft DSA also highlights user choice in shaping RS and in, at the very least, having a choice between a personalised and non-personalised system.

User control of algorithms has also attracted attention in the social media industry. Twitter announced the research project "blue sky" that aims to build an "app store for (…) algorithms".[83] The goal is decentralisation of algorithms used by social media that allows users to control the algorithms shaping the information they see. One element that goes beyond current approaches in decentralised networks like Mastodon is the idea of creating choice for content moderation algorithms. In a conversation with investors, Twitter CEO Jack Dorsey framed the idea as follows:

> *The problem of discovery around content is one that is easiest when it is centralized, and that's how we've operated for almost the past 15 years. But even that has some potential to shift. And one of the things we brought up last year in our Senate testimonies … is giving more people choice around what relevance algorithms they're using for ranking algorithms you're using. You can imagine a more market-driven and marketplace approach to algorithms. And that is something that not only we can host but we can participate in.*[84]

This is one specific example of how user control could be implemented for content moderation by creating a market for content moderation algorithms which would give users a choice between different algorithms.

---

[79] Jin, Cardoso and Verbert (n 78) 38.

[80] Michael D Ekstrand and others, 'Letting Users Choose Recommender Algorithms: An Experimental Study', *Proceedings of the 9th ACM Conference on Recommender Systems* (ACM 2015) <https://dl.acm.org/doi/10.1145/2792838.2800195> accessed 5 March 2021.

[81] F Maxwell Harper and others, 'Putting Users in Control of Their Recommendations', *Proceedings of the 9th ACM Conference on Recommender Systems* (ACM 2015) 8 <https://dl.acm.org/doi/10.1145/2792838.2800179> accessed 5 March 2021.

[82] Jin, Cardoso and Verbert (n 78) 40.

[83] Jacob Kastenakes, 'Twitter's Jack Dorsey Wants to Build an App Store for Social Media Algorithms' (*The Verge* 9 February 2021) <https://www.theverge.com/2021/2/9/22275441/jack-dorsey-decentralized-app-store-algorithms>.

[84] 'Twitter, Inc.'s (TWTR) CEO Jack Dorsey on Q4 2020 Results - Earnings Call Transcript' (*SeekingAlpha*) <https://seekingalpha.com/article/4404806-twitter-inc-s-twtr-ceo-jack-dorsey-on-q4-2020-results-earnings-call-transcript> accessed 16 April 2021. This refers back to an idea of Stephen Wolfram to give users a choice in content moderation.

### *d. A new freedom of association*

As mentioned above, collaborative filtering RS techniques are based on grouping 'similar users' together to drive the predictive power of the models. Through profiling, RS classify and group users together. While measures of user control would influence or change the ways in which people are profiled or the way content is targeted to them, it is also possible to give users power to influence how user groups are formed or, at the very least, how they themselves are grouped. This could be done, for example, through RS allowing individuals to directly associate themselves with a certain group.[85] Certain RS are already exploring this opportunity with regard to gender. A famous fashion RS explores the possibility of allowing users to be more fluid with their gender for the purposes of recommending items to them. Instead of asking whether the user is male, female or something else, they want to know whether somebody would feel male, female or something else.[86] Generalising this idea would mean that the possibility for users to choose a certain group, category, or label they could be characterised with could be a design feature of RS. This would transgress the notion of data protection and its focus on data being correct and up-to-date. It would allow users to associate themselves with groups depending on their will at particular times. This would function as a loose reminder of the freedom of association as a human right in the sense that the freedom of association also encompasses the right to be part or not to be part of a group.[87]

One might object to such a design feature with the argument that it might harm the accuracy and the fit of the respective recommendation. There might be also further burdens to the optimization of the respective system given that the person choosing the group might not share many of its attributes. Yet there are a number of potential autonomy-enhancing benefits of such an approach. Firstly, certain circumstances may warrant such a feature. This would be cases in which certain individuals have a strong and legitimate interest not to be categorised in a rigid manner, but also where there are

---

[85] One could also think about interactive possibilities of recommending things amongst users. Bart P Knijnenburg, Saadhika Sivakumar and Daricia Wilkinson, 'Recommender Systems for Self-Actualization', *Proceedings of the 10th ACM Conference on Recommender Systems* (ACM 2016) 12 <https://dl.acm.org/doi/10.1145/2959100.2959189> accessed 29 July 2021.

[86] This information is based on an expert interview.

[87] Christian Tomuschat, 'Freedom of Association' in Ronald J St Macdonald (ed), *The European System for the Protection of Human Rights* (Nijhoff 1993); Jürgen Bröhmer, 'Kapitel 19: Versammlungs- und Vereinigungsfreiheit' in Oliver Dörr, Rainer Grote and Thilo Marauhn (eds), *EMRK/GG. 2: Kapitel 20 - 33, Register* (2. Aufl, Mohr Siebeck 2013). Of course, the freedom of association as a human right requires some stability of the respective group which would not be the case.

no potential negative consequences from a recommendation based on the user's self-determined grouping. For example, in the case of gender, a fashion recommendation would pose no harm regardless of gender specified, however health recommendations may be based on research that is biologically gender-specific. Secondly, a freedom of association would also allow for the intuitive self-determination of users who might not have expertise how the system works but might gain some experience about the association with certain groups, which produces the best outcomes for them. A limitation of this approach is that it may be specific to RS that rely on a communicative relationship between a human and a computer in which the ultimate decision rests with the human being.

### *e. Inter-subjective autonomy*

In all the above-mentioned cases, design features address autonomy at the level of an individual user. However, this misses the importance of group or collective autonomy. In the process of profiling and grouping users, an influential decision could be made about which feature similarities are relevant, and thus become a group and which do not form a group of their own. Thus, profiling constructs groups of users but leaves other potential groups unconstructed. This provokes the important design question of whether inter-subjective autonomy can also be exercised through the design of RS. Is there a possibility for groups to determine themselves? This line of thinking can draw upon different ideas such as pluralism or other conceptions focusing on the interests of developing states such as post-colonial computing.[88] Inter-subjective autonomy requires design features for groups to influence the design of RS, and at the very least to establish their existence in the 'eyes' of an algorithm. A first step would be to define certain classes that are not present if categories like gender or ethnicity are narrowly constructed. As a next step, if a group is constructed and this group can express its preferences, it might be possible to allow this group to influence the respective recommender system in the ways described above.

What is clear from the foregoing discussion is that the way RS are designed can shape the impact they have on individual and collective autonomy. Moreover, we have shown that recent legislative initiatives show indications for the requirement of designers and deployers to consider autonomy in the process of creating RS. We have also introduced a number of technical design

---

[88] Lilly Irani and others, 'Postcolonial Computing: A Lens on Design and Development', *Proceedings of the 28th International Conference on Human factors in Computing Systems - CHI '10* (ACM Press 2010) <http://portal.acm.org/citation.cfm?doid=1753326.1753522> accessed 29 July 2021.

measures that could either help minimise the intentional, human-designed impact of RS on autonomy or can help maximise autonomy and empower RS users. Nevertheless, we also highlighted that there are more decisions for regulators to make and clarify, including what a potential autonomy-by-design obligation would entail – simply accounting for and mitigating the impact RS have on autonomy or rather actively considering how to use technology in a manner that empowers individuals to pursue their life paths. Moreover, further research will be necessary to understand how to define and assess degrees of human autonomy and independence in human-computer interaction, as well as whether and how different contexts of RS application justify a differential approach to implementing autonomy-by-design in practice. Regardless, it is clear that design must be one of the regulatory pathways to truly safeguarding autonomy.

## B. Input: Governance of personal data

It would be unfair to say that individuals currently have no recourse to control how they are 'seen' and profiled, including by RS. Privacy and data protection legislation have increasingly been adopted all over the world. This is relevant to autonomy because through inferences, grouping, and classification, RS can also interfere in the personal identity experience by, for instance, classifying a profile in a manner not corresponding to the features or categories with which the user self-identifies.[89] Moreover, the use of automated inferences[90] can reinforce biases, stereotypes, and stigmas, even without people's awareness. These inferences can significantly affect people's privacy, identity, and self-determination.[91] In that context, data protection regulation can be a tool to govern the development and use of RS. The question of how legal rights can empower individuals to shape the data input of RS in a manner that safeguards individual autonomy is further explored below.

---

[89] Milano, Taddeo and Floridi (n 2) 962.

[90] In Europe, there is still no consensus on the classification of inferences made by automated systems based on information about people. For the Article 29 Working Party, it would be classified as personal data and, then, protected under GDPR, but the Court of Justice of the European Union disagrees with such an approach. (Joined Cases C-141/12 & C-372/12, *YS, M and S v Minister voor Immigratie, Integratie en Asiel* [2014] OJ C 315); Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) Columbia Business Law Review 494; Article 29 Data Protection Working Party (n 68) 5.

[91] Wachter and Mittelstadt (n 90) 513.

## i. State of the art in European data protection law

The EU has the GDPR which does not regulate RS specifically, but rather the collection and use of any personal data,[92] including for profiling and automated decision-making. The GDPR[93] plays an important role in safeguarding individual autonomy because it strengthens individual control over personal data.[94] It is addressed at public authorities and private actors alike. It, therefore, fulfils the obligations of states to respect and protect human rights in society. The regulation embodies the principle of informational self-determination, setting specific obligations for data controllers while protecting and empowering individuals.[95] There are a number of key features of the current data protection legislation in this regard.

### *a. Consent*

In order to ensure greater individual self-management of data,[96] consent, as an expression of free choice, self-determination and autonomy, plays an important function in data protection.[97] It is essential to the exercise of individual control over personal data.[98] The GDPR establishes explicit, free and *informed* consent as one of the lawful bases of art 6, permitting the processing of personal data and legitimizing algorithmic processing of personal data.[99] Although there are other legal bases in the regulation, the data

---

[92] Personal data means "any information relating to an identified or identifiable natural person ('data subject')". Art. 4 (1) Regulation 2016/679 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.

[93] Although this study was mostly based on the European scenario, it is worth mentioning that the GDPR was a robust data protection regulation that inspired many other countries, not only on the drafting of their data protection bills but also stimulating a higher level of enforcement and compatibility with the European guideline. In Latin America, for instance, this was the case of Brazil, which built its Data Protection Regulation mirroring the GDPR envisioning an enhanced privacy culture internally, and the possibility of enabling lawful international transfers, and stimulating companies to uniformize its policies on an international level with a GDPR standard.

[94] Tatiana Shulga-Morskaya, 'Protection of Personal Data through Implementation of the Right to Informational Self-Determination: Identifying Opportunities and Pitfalls' (2019) <https://www.giga-net.org/2019symposiumPapers/17_Shulga-Moskaya_PROTECTION-OF-PERSONAL-DATA.pdf> accessed 30 March 2021.

[95] ibid.

[96] Daniel J Solove, 'Privacy Self-Management and the Consent Dilemma' (2013) 126 Harvard Law Review 1880.

[97] Bart W Schermer, Bart Custers and Simone van der Hof, 'The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection' (2014) Ethics and Information Technology <http://link.springer.com/10.1007/s10676-014-9343-8> accessed 5 March 2021.

[98] Solove (n 96).

[99] Bruno Ricardo Bioni, *Proteção de Dados Pessoais - A Função e Os Limites Do Consentimento* (2nd edn, Forense 2019).

subjects' consent plays a central role in the law regarding autonomy, since it allows genuine and informed individual control over an individual's data.[100] When consent is obtained in full compliance with the conditions imposed by the GDPR, it is an effective tool to ensure users' control whether or not personal data concerning them will be processed,[101] which enables autonomy. Consent can be an especially valuable and necessary safeguard in the context of intrusive activities such as in the case of decision-making based solely on automated processing that, in other circumstances, would be prohibited by the law.[102] The Article 29 Working Party, a former expert body providing authoritative interpretations of European data protection law, furthermore suggested that in most of the cases of algorithmic data processing, such as in RS, which affect individual and collective autonomy, focus should be on getting the user's consent.[103] Upon closer examination, there are a number of requirements for ensuring consent actually safeguards individual autonomy, which are not always easily met in practice.

According to art 4(11) of the GDPR, valid consent must be freely given, specific, informed, and unambiguous, through a clear statement of affirmative action that indicates the data subject's wishes and agreement to the processing of their personal data.[104] It is vital that consent is informed, meaning that individuals are provided sufficient information to understand what they are asked to agree to, including what data would be processed, by whom and

---

[100] Article 29 Data Protection Working Party, 'Guidelines on Consent under Regulation 2016/679' (2018) WP259 rev.01 <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051> accessed 30 March 2021.

[101] European Data Protection Board, 'Guidelines 05/2020 on Consent under Regulation 2016/679, Version 1.1' (4 May 2020) 5 <https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf> accessed 30 March 2021.

[102] ibid 18.

[103] Article 29 Data Protection Working Party (n 100) 47.

[104] ibid 5; European Data Protection Board (n 101) 7–18. To be considered valid in the terms of the regulation, consent must be simultaneously:

(i) freely given – meaning a real choice and control for data subjects. If the user feels compelled to consent or will endure negative effects by not consenting, consent will not be considered informed, thus, invalid. Also, consent is not free when there is not an option to refuse or withdraw consent without detriments – it must not be considered a condition;

(ii) specific – users' consent must be directed to one or more specific purposes, giving them a choice in relation to each of these purposes (granularity). This enables control and transparency;

(iii) informed – meaning that the controller, before obtaining consent, must provide users with enough information to ensure informed decision making. For example, informing users about what they are agreeing to and how to exercise their right to withdraw, if necessary;

(iv) unambiguous indication of the data subjects' wishes to authorize the processing of their data – it must be given through an active motion or declaration by the user, making clear and obvious that they accepted and understood the terms.

for what purpose.[105] This is linked to a right to receive information, necessary for the validity of consent.[106] The importance of information is further discussed also in Section III.C of this paper.

Consent must also be unambiguous and explicit in that, as clarified in recital 32 of the GDPR, silence, pre-ticked boxes or inactivity should not be accepted as consent. Despite the non-binding status of the GDPR's recitals, this provision reinforces the voluntary and non-mandatory nature of consent, as it must be actively given in order to maintain the individual's control over data.[107] The inclusion of the obligation to inform users regarding the possibility to withdraw consent confirms that it is reversible, which puts a degree of control on the side of RSs' users.[108] The e-Privacy Directive[109] also requires informed and prior consent for all except the necessary technical cookies on websites, rejecting opt-out mechanisms for all other cases, but rather requiring explicit user action to indicate consent. This is positive for the exercise of individual autonomy since the user must decide and express their active choice for use of tracking technology, in a measure of opt-in.[110] In the same vein, the European Court of Justice decided, in the case of *Planet 49*, that pre-selected checkboxes are insufficient to obtain valid consent for placing cookies on users 'systems, as it does not constitute an unambiguous indication of their wishes.[111]

Finally, consent must be free in that the data subject is offered an effective control over his data and, in the context of RS, has a genuine choice with regard to accepting or declining (without detriments) the terms of the service.[112] However, digital platforms which are the largest users of RS, fail to provide real alternatives for consent, instead presenting the users with a 'take it or leave it' choice.[113] This undermines the requirement for 'free' consent,

---

[105] GDPR, arts 13, 14.

[106] European Data Protection Board (n 101) 94.

[107] Iris van Ooijen and Helena U Vrabec, 'Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective' (2019) 42 Journal of Consumer Policy 91, 100.

[108] ibid 7.

[109] Directive 2002/58/EC of the European Parliament and of the Council of July 12, 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002] OJ L 201.

[110] Martino Trevisan and others, '4 Years of EU Cookie Law: Results and Lessons Learned' (2019) 2019 Proceedings on Privacy Enhancing Technologies 126, 138.

[111] Case C-673/17 *Verbraucherzentrale Bundesverband eV v Planet49 GmbH* [2019] ECLI:EU:C:2019:801.

[112] European Data Protection Board (n 101).

[113] Varshney (n 39); Helen Nissenbaum, 'A Contextual Approach to Privacy Online' (2011) 140 Daedalus 32; Mariarosaria Taddeo and Luciano Floridi, 'The Debate on the Moral

thus affecting user autonomy and agency. In this case, the users' control is illusory, and consent could be questioned as a basis for the processing of personal data that could be perceived as unlawful.[114]

### *b. Responsibilities of data controllers and processors*

Apart from empowering users by giving them control over their personal data through consent, the GDPR also enhances autonomy by balancing the regulatory burden across the different key actors of the data network processing, especially in terms of the need for compliance of the obligations related to the principles, accountability and data subject's rights protection.[115] Data controllers and processors are namely those responsible for processing personal data in compliance with a number of legal principles that seek to establish a general framework that balances the interests of individuals with the controllers' and processors'. For example, art 7 (1) and recital 42 of the GDPR place the burden of demonstrating the compliance with the requirements of valid consent on data controllers. Moreover, even with the person's consent, both data controller and processor still must comply with data protection principles of GDPR's art 5(1) and (2),[116] which are: (a) lawfulness, fairness and transparency; (b) purpose limitation; (c) data minimisation; (d) accuracy; (e) storage limitation; (f) integrity and confidentiality; (2) accountability. Also, the processing must be legitimized by one of the legal bases presented in art 6 (1), attached to specific purposes, and the personal data involved has to be accurate, updated, adequate, relevant and strictly limited to what is necessary for this purpose that was accepted by the user in the moment of consent. Thus, for instance, even in the case of personal data processing in RS based on consent, this would not legitimize the collection of excessive data in relation to a particular purpose.[117] The GDPR requires even stronger compliance when the processing involves "special categories of personal data,"[118] which demands a second layer of legal basis, which are presented in art 9(2)of the GDPR. Therefore, if the RS is based on the processing of sensitive data, it would imply a higher data processing risk which leads

Responsibilities of Online Service Providers' (2016) 22 Science and Engineering Ethics 1575.

[114] European Data Protection Board (n 101) 5.

[115] Alexandra Giannopoulou, 'Algorithmic Systems: The Consent Is in the Detail?' (2020) 9 Internet Policy Review <https://policyreview.info/node/1452> accessed 5 March 2021.

[116] Article 29 Data Protection Working Party (n 100) 13.

[117] Article 29 Data Protection Working Party (n 100).

[118] According to art 9(1) GDPR, special categories of personal data are related to: "Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited".

to the necessity of enhanced compliance and stronger safeguards. Thus, the GDPR imposes significant obligations and requirements on data controllers in order to preserve and strengthen human autonomy.

### *c. Data protection rights for empowering individuals*

In the monitored, surveilled and data-driven society, the safeguard of individual and collective autonomy online must also rely on a data protection subject's rights, as they derive logically from the aforementioned data protection principles. These rights are intended to empower users to control what happens to their data. Following the GDPR's principle of accountability, the key actors of RS's data processing will need to demonstrate their compliance with the regulation in general, and specifically that they can provide data subjects' rights through effective mechanisms and internal processes.[119] In that sense, GDPR embodies important data subject's rights, actionable against the controller during all the steps of processing. This includes the moment of creating the profile and also when making the automated decision about the user, based on his profile, with the purpose to recommend items. Even where a user consents to their personal data being processed, for example, the rights of arts 15-20 of the GDPR are still applicable,[120] which enable users to, *inter alia*, supervise the processing of their data and, when necessary, make updates, ask for additional information or even object to the processing of their data.[121]

Among the key rights is the right to be informed. As a consequence of the principle of transparency (art 5(1)(a) and recital 60 GDPR), RS's controllers must proactively inform data subjects about their rights, the existence of data processing and all information related to it, including its purposes, besides a clear, meaningful and understandable explanation of how profile and RS techniques work,[122] which is provided in arts 13 and 14 of GDPR. These provisions encompass the right of the data subjects to receive information from the controller, who has the legal obligation to inform them, even without request. According to art 12, the controller must freely provide information to the data subject, in a concise, transparent, accessible, and

---

[119] Álvaro Tejeda-Lorente and others, 'Adapting Recommender Systems to the New Data Privacy Regulations' in Hamido Fujita and Enrique Herrera-Viedma (eds), *Volume 303: New Trends in Intelligent Software Methodologies, Tools and Techniques* (IOS Press eBooks 2018).

[120] Article 29 Data Protection Working Party (n 100) 30.

[121] The right to object (GDPR, art 21) does not apply when consent is the legal basis for the processing. However, a similar outcome is possible, since people can withdraw consent at any time, as easy as giving it and without detriments; ibid 21, 22, 30.

[122] Tejeda-Lorente and others (n 119) 16–17.

easy way, and also facilitate the exercise of their rights under arts 15 to 22, which are: right to access, to rectification, to erasure, to processing restriction, to data portability, to object to processing, and not to be subject to a decision based solely on automated processing.[123]

The right to access, under art 15 and recital 63, reinforces the right to information of the previous articles, as it allows individuals to actively request information from the controller. In this sense, people may require confirmation of the existence of personal data processing concerning them and also the presence of automated decision-making for recommendation, which can be used for profiling. Where that is the case, the subject must be able to access his personal data, all information related to its processing and also meaningful information about the logic involved in the automated profiling techniques. This access may also enable the exercise of other important rights (depending on the situation and legal basis), such as rectification to update or amend inaccuracies (art 16 GDPR), erasure (art 17 GDPR), restriction of processing (art 18 GDPR) and object (art 21 GDPR).

In terms of individual self-determination, as an expression of autonomy, the right to information and access to personal data is a powerful instrument, since it provides users with the fundamental basis to understand the processing of their data, the RS's techniques and, thus, to make informed decisions accordingly.[124] In some circumstances, these rights may give people greater knowledge about the logic involved in the recommendations they receive, which allows them to exercise other rights, for example rights of rectification, erasure and portability.[125] According to the European Data Protection Supervisor, the right to data portability would allow people to use data for their own purposes and exercise their option to change information service providers.[126] Thus, it is understood as an expression of individual autonomy and empowerment, as it enables individuals to access and then transfer their personal data from one platform to another, without detriments.[127] This also

[123] Shulga-Morskaya (n 94) 8. In the Indian context, most of the above listed rights can be found under the data protection Bill, for a detailed study of this, see Pallavi Bedi, 'Comparison of the Personal Data Protection Bill with the General Data Protection Regulation and the California Consumer Protection Act' (The Centre for Internet & Society 2020) <https://cis-india.org/internet-governance/blog/comparison-of-the-personal-data-protection-bill-with-the-general-data-protection-regulation-and-the-california-consumer-protection-act-2>.

[124] van Ooijen and Vrabec (n 107) 94.

[125] ibid 102.

[126] European Data Protection Supervisor, 'EDPS Recommendations on the EU's Options for Data Protection Reform' (European Data Protection Supervisor 2015) 2015/C 301/01 7 <https://eur-lex.europa.eu/legal-content/GA/TXT/?uri=CELEX:52015XX0912(01)> accessed 30 March 2021; van Ooijen and Vrabec (n 107) 102.

[127] van Ooijen and Vrabec (n 107) 102.

serves to enhance competition between service providers and could make it an important competitive feature, insofar as how individuals assess and perceive the digital services they can choose from and the adequacy of the treatment of their data. This is confirmed by GDPR's recital 68 that sustains the idea of data portability rights as a form of strengthening users' control over their own data, where the processing happens by automated means. By setting these principles and rights, the GDPR effectively safeguards the power of individuals to exercise their autonomy by managing their data in line with their preferences.[128] Moreover, compliance with the GDPR also ensures companies and governments respect and fulfil the individual's rights and freedoms.

### ii. Ways to further enhance autonomy

#### *a. Truly informed exercise of rights*

Despite the guarantees and protections afforded by the GDPR to individuals to empower them in the control of their personal data, some of its provisions are still difficult or inconvenient for controllers to abide by. One such example is the issue of truly informed consent. In practice, the consent often incorporated in the privacy policies of large platforms can be perceived as ineffective. Instead of empowering users, it operates as a way to legitimize business models of the information economy to "adapt" to the GDPR rules.[129] This scenario may deprive individuals' agency since the consent given by the user is rarely informed in an adequate way, but rather a condition to access the service.[130] Given the impossibility of negotiating the terms of service, people tend to focus on the immediate benefit (access to a product or service online), to the detriment of the possible long-term harm to their privacy, which can reinforce the loss of control over their data.[131] This is especially the case when these platforms embody algorithm-based profiling, nudging and even manipulation, as is the case in RS.

Another reason behind the difficulty of attaining truly informed consent is the challenges for individuals to actually understand how their personal data is processed, and to what end, by AI techniques. These techniques may

---

[128] Paul Bernal, *Internet Privacy Rights: Rights to Protect Autonomy* (Cambridge University Press 2014) 14.

[129] Izabella Alves Jorge Bittencourt and Évelyn Vieira Gomes, '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 Fabrício Polido, Lucas Anjos and Luíza Brandão (eds), *Políticas, Internet e Sociedad* (Instituto de Referência em Internet e Sociedade 2019).

[130] Varshney (n 39).

[131] Bioni (n 99).

be technically opaque and unpredictable, considered "black boxes" or may be protected by trade secrecy. Both these types of protections are further discussed in Section III.C and may hinder the right to information and measures of explanation and transparency that are essential to the effective exercise of autonomy through consent and the data subject's rights, mainly those rights associated with information and access. Sophisticated AI algorithms used in RS are not easily explainable to data subjects and sometimes even for controllers, as the technology may operate in unpredictable ways.[132] Therefore, individuals are placed in a situation of informational, technical and economic asymmetry, where the lack of foresight makes it difficult to ensure informed consent and, consequently, autonomy.[133] Moreover, individuals may also be confronted with the controllers' interests related to intellectual property and industrial secrecy[134] which further obfuscates the information necessary for them to exercise their rights in an informed way.

A possible new e-Privacy Regulation, still in the draft and discussion phase, may help address some challenges around making consent actionable without overwhelming users.[135] On January 5, 2021, despite some criticism,[136] the Portuguese presidency of the Council of the European Union published the 14th draft of the regulation, which is simpler and aligned with the GDPR[137] and would replace the current e-Privacy Directive. In line with

---

[132] van Ooijen and Vrabec (n 107) 96.

[133] Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) 16 IEEE Security & Privacy 46.

[134] Wachter and Mittelstadt (n 90) 498–499.

[135] An interesting development in the Indian context is the role of consent managers as proposed under the Personal Data Protection Bill 2019. Consent Managers as envisaged under the Bill act as data fiduciaries that enable data principals to delegate the exercise of their agency. For a detailed study see Samraat Basu and Siddharth Sonkar, 'Regulating Consent Managers in India: Towards Transparency and Trust in the Digital Economy' (Oxford Law Faculty, 1 April 2020) <https://www.law.ox.ac.uk/business-law-blog/blog/2020/04/regulating-consent-managers-india-towards-transparency-and-trust> accessed 15 January 2022.

[136] The German Federal Commissioner for Data Protection and Freedom of Information (BfDI), Professor Ulrich Kelber, criticised the last draft of the e-Privacy Regulation, as he considered it a risk to data protection and privacy. According to the Commissioner's interpretation of the document, the draft would allow the use of cookie walls, which is considered a hindrance in individuals' protection on the internet. Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (BfDI), 'BfDI Kritisiert Position Des Rats Zur EPrivacy-Verordnung' (10 February 2021) <https://www.bfdi.bund.de/DE/Infothek/Pressemitteilungen/2021/03_Ratsposition-ePrivacy-VO.html> accessed 30 March 2021.

[137] Dan Cooper and Anna Oberschelp de Meneses, 'Council of the EU Released a (New) Draft of the ePrivacy Regulation' (Inside Privacy, 6 January 2021) <https://www.insideprivacy.com/eu-data-protection/council-of-the-eu-released-a-new-draft-of-the-eprivacy-regulation/> accessed 30 March 2021; Presidency, 'Proposal for a Regulation of the European Parliament and of the Council Concerning the Respect for Private Life and the Protection of Personal Data in Electronic Communications and Repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)' (Council of the European Union,

the idea of giving greater control to users and thus guaranteeing their autonomy, the draft focuses on consent for the treatment of electronic communication data, a wider category of data than personal data. Consent is required whether for the processing of the content of electronic communication data, metadata or information from the terminal equipment of the user. Although the draft refers to the GDPR's definition of consent, the document attempts to address some of the problems associated with that consent, such as overloading requests or mandatory consent to access certain services. Among the possible solutions raised in the draft is the possibility of implementing technical means in electronic communications software to allow specific and informed consent through transparent and easy settings for users. Thus, it would allow end users, in a transparent and friendly manner, to manage consent for the storage and access to data stored on their terminal equipment, easily configuring, changing, and withdrawing consent at any time.[138]

### *b. Greater control over inferred data*

Greater transparency and information may be especially important in the case of inferred data where individuals do not directly provide information about themselves, but instead assumptions about them are made on the basis of other data and their behaviour. More control of individuals over the way they are viewed and the assumptions made about them may be desirable. RS may use data to create inferences about a person on the basis of which they make recommendations that may interfere in their behaviour, thus giving rise to a risk to individual reputation, privacy, self-determination and autonomy. Even though the current GDPR framework provides for detailed governance of personal data that could be input into RS, it still lacks protection against how data is subsequently evaluated.[139] Thus, we still face accountability gaps in the GDPR; for instance, for data processing related to inferences that may be inaccurate, biased, and even sensitive.[140] This could especially be a problem in situations where inferences relate to data that would otherwise be considered sensitive, e.g. gender, sexual orientation or religious beliefs. Such inferences could, moreover, be based on anonymous or non-personal data – another type of data not covered by the GDPR but that could nevertheless pose risks to data subjects.[141] To cover these gaps, the GDPR should include not only personal data, but also the accuracy of decision-making processes

2021) 5008/21 <https://data.consilium.europa.eu/doc/document/ST-5008-2021-INIT/en/pdf> accessed 30 March 2021.

138 Presidency (n 137) 38.

139 Wachter and Mittelstadt (n 90) 620.

140 ibid 613.

141 ibid 615–618.

and the assessment of the reasonableness of inferential analysis carried out by algorithms.[142]

The draft e-Privacy Regulation also demonstrates that safeguards are necessary for data beyond personal data. In its current version, this regulation would be broader than the GDPR, since it is not limited to the processing of personal data; rather, it is applicable to electronic communications data.[143] This incorporates both the content and the metadata of these communications, which may include sensitive information, even if not classified as personal data, such as website visited, geographical location, time, date and duration of some website's use.[144] This information may be used in RS's data processing and monitoring techniques in order to create users, profiles and would now be protected.

Some have suggested the existence of a new right to reasonable inferences, which would also provide for the associated right to challenge unreasonable high-risk assumptions.[145] This possibility would enable individuals to object to certain inferences or the irrelevance, lack of confidence or inaccuracy of data used to create those inferences, going beyond the current right of individuals to rectify their personal data by correcting inaccurate data. As a result, these practices would empower individuals to exercise control over their data, reinforcing the right to access and rectification, while also complementing the right to challenge solely automated decisions, including profiling.[146] This could also help implement in practice the above-mentioned freedom of association that could allow individuals to freely choose which groups or labels they are or are not associated with.

### *c. Impact assessments going beyond data protection*

Another way of enhancing individual autonomy through data protection is by providing actionable tools for those handling personal data to appropriately and lawfully handle data. One option is for providers of RS to implement data protection risk and impact assessments, in accordance with art 35 of the GDPR and as a best practice. The draft e-Privacy Regulation also establishes obligations or advice for the implementation of impact assessments, referring to the already existing art 35 of the GDPR. Even though the legal provisions of the GDPR mainly deal with issues related to privacy, the risks that RS give effect to make it recommendable to go further. RS

[142] ibid 615.
[143] Cooper and Oberschelp de Meneses (n 137); Presidency (n 137).
[144] Presidency (n 137) 11.
[145] Wachter and Mittelstadt (n 90) 619.
[146] ibid 619–620.

designers or implementers could implement algorithm audits and algorithmic impact assessments to map the RS risks related to legal compliance and ethical guidelines, human rights, especially autonomy, but also fairness (bias audits), non-discrimination, due process and ensuring the public oversight.[147] Audits can help secure compliance with existing legal and ethical standards, while algorithmic impact assessment, including algorithmic risk assessment and impact evaluation, may help assess possible societal impacts on the autonomy of RS before and during its implementation in real life.[148] By acting to alleviate any shortcomings identified, risk assessments and audits, particularly through agile design decisions, could serve as valuable governance tools to help RS creators strengthen autonomy and self-determination of their users in practice.

### *d. Recent legislative initiatives*

The recent legislative initiatives of, i.e. the aforementioned DSA and the AI Act, also have a role to play in further developing data governance frameworks and data rights of individuals. As said before, the DSA devoted considerable attention to RS, especially in art 29 and recital 62. This provision addresses "very large platforms" that use RS, requiring them to set their terms and conditions in a clear, accessible, and comprehensible manner to inform users of the RS and, where possible, inform them of options to influence the recommendations. This draft's obligation would empower users through information, being a step beyond the focus of the GDPR on users' ability to exercise control over their data.[149] Although it is noteworthy that the DSA is a first initiative to specifically address RS, the proposal is only applicable to large online platforms and is still vague. It does not explain the possible options that users should have, in terms of influencing recommendations sent to them nor a way to align this with other fundamental rights. For the regulation to effectively give users control over their data in RS, the draft could, for example, require the implementation of democratic and fairer recommender algorithms or enable users to effectively choose between different recommendation algorithms, including from third parties.[150]

---

[147] Ada Lovelace Institute and Data Kind UK, 'Examining the Black Box: Tools for Assessing Algorithmic Systems' (Ada Lovelace Institute 2020) 3 <https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/> accessed 6 March 2021.

[148] ibid.

[149] Natali Helberger and others, 'Regulation of News Recommenders in the Digital Services Act: Empowering David against the Very Large Online Goliath' (Internet Policy Review, 26 February 2021) <https://policyreview.info/articles/news/regulation-news-recommenders-digital-services-act-empowering-david-against-very-large> accessed 30 March 2021.

[150] ibid.

In addition, the European Commission has expressed awareness of the need to address the specific challenges that AI systems may create.[151] Thus, the recently proposed AI Act aims to foster the development of an ecosystem of trust in AI in Europe.[152] RS would fall within the definition of AI within the draft Act[153] and depending on the scope of the RS, it could be classified in one of the four levels of risk created by the AI Act. The proposal follows a risk-based approach, defining the possible uses of AI according to whether they create an unacceptable, high, limited or minimal risk to people's security and fundamental rights. According to recital 14, depending on the intensity and the scope of the risks of AI systems, some systems may be prohibited. Indeed, as was mentioned in the introduction, AI systems developed with a "*significant potential to manipulate persons through subliminal techniques (…) or exploit vulnerabilities (…) in order to materially distort their behaviour in a manner that is likely to cause them or another person psychological or physical harm*" would be prohibited by the AI Act, as they are considered a threat to safety, livelihoods and rights of people. Where manipulative or exploitative practices, facilitated by AI, are not prohibited, the draft proposal points to other potential legal safeguards to ensure individuals are sufficiently informed and can freely choose whether or not to be subjected to profiling that could affect their behaviour - data protection law, consumer protection, or digital services legislation.[154] The latter legislative body, particularly, may soon be modernised in the EU through the DSA. Nevertheless, as we have seen so far, there are gaps in some of these mentioned legal frameworks in terms of the protection they offer against manipulation or influence through RS.

Where the requirements for prohibition are not met, high-risk AI applications are subject to strict requirements of risk management and reporting on data governance, transparency, human oversight, accuracy, robustness, and

---

[151] European Commission, 'Regulatory Framework on AI | Shaping Europe's Digital Future' (*European Commission*, 1 July 2021) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 29 July 2021.

[152] Commission, "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" (n 25).

[153] Recital 6 of the AI Act proposal. Further, the definition is on art 3 (1) of the AI Act: "artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with".

[154] Commission, "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" (n 25) 13.

cybersecurity.[155] Depending on the purposes, the modalities of use and the function performed by the RS, it could be classified as high-risk, as it may create threats to peoples' health, safety or fundamental rights. The list of high-risk AI systems is focused on specific use cases in the fields of biometric identification, management of critical infrastructure, education, employment, access to public services or essential private services, law enforcement, border control, or the administration of justice.[156] The list specifically focuses on systems used to make decisions, and as such, it is an open question whether RS would fall within that scope, given their 'advisory' role in shaping human decision-making. Nevertheless, it is conceivable to imagine RS used to provide rankings that could then be used to prioritise needs and direct resources or workflow, for example, in the context of education or employment.

Where the requirements for meeting the high-risk threshold are also not met, AI systems are subject to significantly fewer obligations; yet, they are important for safeguarding individual autonomy. In such cases, a RS would need to comply with certain transparency obligations, such as the delivery of information to users that they are interacting with an AI, in order to allow their informed decision.[157] The multiple new requirements arising from the AI Act would overall have an impact on the way RS are designed, created, and maintained, and create an incentive to RS's providers to promote compliance by design in the case of RS.[158]

To sum up, the GDPR and other digital technology-related regulations like the draft DSA, the draft AI Act, and the draft e-Privacy Regulation, try to develop a stronger culture of informational self-determination associated with data protection in the context of RS and are important to ensure enhanced control, through the effective exercise of data subjects' rights and lawful consent. For example, after the GDPR adoption and the last amendments in the current e-Privacy Directive, European consumers encountered significantly less unconditional usage of persistent cookies when using the Internet and its services.[159] Already in the early days of the GDPR, in 2018,

---

[155] AI Act, ch 2.

[156] AI Act, annex III.

[157] AI Act, art 52.

[158] Friederike Reinhold and Angela Müller, 'AlgorithmWatch's Response to the European Commission's Proposed Regulation on Artificial Intelligence – A Major Step with Major Gaps' (AlgorithmWatch, 22 April 2021) <https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021> accessed 29 July 2021.

[159] Adrian Dabrowski and others, 'Measuring Cookies and Web Privacy in a Post-GDPR World' in David Choffnes and Marinho Barcellos (eds), *Passive and Active Measurement*, vol 11419 (Springer International Publishing 2019).

transparency measures increased 4.9%; more websites had privacy policies and informed their users about cookies practices, data subjects' rights and the legal basis for processing of personal data.[160] Nevertheless, current privacy-related regulations alone might not be considered sufficient to guarantee an adequate level of autonomy on matters specific to autonomy and RS. the strengthening of users' protection is necessary and examples exist, such as the changes brought by the new draft of the e-Privacy Regulation, the implementation of limits, obligations and requirements for AI systems in the AI Act (that would be enforced and supervised by the European Artificial Intelligence Board and national authorities within the Member States), and the creation of new rights, such as the right to reasonable inferences. All of these are important steps towards safeguarding autonomy. In parallel, solutions to embed autonomy in the design of RS through the prism of data governance, as well as solutions coming proactively from the private sector,[161] society and technology (for example, the law-by-design approach and its implementation through audits and risk assessments) should be considered and fostered simultaneously.

## C. Output: Communication and Transparency

A final aspect where regulation can play a key role in enhancing individual autonomy is through transparency. It is a widely supported principle in AI ethics frameworks[162] and is one of the five OECD AI Principles,[163] endorsed by the G20 countries,[164] as well as a requirement in the EU's Trustworthy AI

---

[160] Martin Degeling and others, 'We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy' (2019) Proceedings 2019 Network and Distributed System Security Symposium <http://arxiv.org/abs/1808.05096> accessed 30 March 2021.

[161] For example, Google announced the intention to remove support for third-party cookies in their browser Chrome and that the company is working on the development of a Privacy Sandbox to build innovations that would protect anonymity while still delivering results for advertisers and publishers. Google made explicit that it will not replace third-party cookies with alternative identifiers to track individuals as they browse across the web nor use them in their products. The company aims to power their products with privacy-preserving APIs, such as Federated Learning of Cohorts (FLoC) which prevent individual tracking while still delivering results for advertisers and publishers; David Temkin, 'Charting a Course towards a More Privacy-First Web' (*Google*, 3 March 2021) <https://blog.google/products/ads-commerce/a-more-privacy-first-web/> accessed 30 March 2021.

[162] Jessica Fjeld and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI' (*Berkman Klein Center for Internet & Society* 2020) 41 <http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420> accessed 6 April 2021.

[163] OECD, 'Recommendation of the Council on Artificial Intelligence' (25 May 2019) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 5 March 2021.

[164] G20, 'G20 Ministerial Statement on Trade and Digital Economy' (8 June 2019) 20 <https://trade.ec.europa.eu/doclib/docs/2019/june/tradoc_157920.pdf> accessed 5 March 2021.

guidelines.[165] Transparency has also been a key feature of the recently proposed EU AI Act and DSA. Importantly, transparency can play a key role in ensuring and safeguarding individual autonomy in the context of RS. Firstly, there is a positive obligation inherent in the principle of respect for autonomy to disclose information necessary to foster autonomous decision-making.[166] Agency is necessary for autonomy and it requires that individuals have sufficient understanding of the environment within which they decide and act, as well as of the meaning and consequences of their decisions.[167] This is reflected, for example, in the notion of informed consent in the GDPR and elaborated in the previous section.[168] Secondly, transparency can safeguard individual freedom by ensuring individuals can control or hold accountable those who may exert control or influence over them, thus further safeguarding self-determination and autonomy. Transparency can also serve to enhance the quality and impact of RS. Explaining to users how an individual recommendation has been made may enhance trust and acceptance by users.[169] Moreover, understanding the model's operation can empower users to adjust their interaction with the RS to produce more desirable recommendations,[170] ultimately improving the service.

Transparency is, however, not a simple matter to regulate. First, what transparency means and what it covers is not a straightforward question. Moreover, transparency may conflict with protecting commercially sensitive or valuable information, as well as private information regarding other users. How to present information so that it is understandable to its target recipients and who they are is a further challenge. Therefore, a nuanced consideration is necessary in order to ensure transparency of RS appropriately safeguards individual autonomy. We define transparency as the availability of information about an actor's workings or performance that allows monitoring or control from others[171] in order to focus on its role as a tool of accountability. In the rest of this section, we explore the current state of

---

[165] High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (2019) <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines> accessed 27 April 2020.

[166] Beauchamp and Childress (n 39) 104.

[167] ibid 102.

[168] GDPR, art 6(1)(a).

[169] Henriette Cramer and others, 'The Effects of Transparency on Trust in and Acceptance of a Content-Based Art Recommender' (2008) 18 User Modeling and User-Adapted Interaction 455.

[170] Donghee Shin, 'User Perceptions of Algorithmic Decisions in the Personalized AI System: Perceptual Evaluation of Fairness, Accountability, Transparency, and Explainability' (2020) 64 Journal of Broadcasting & Electronic Media 541, 549.

[171] Albert Meijer, 'Transparency' in Mark Bovens, Robert E Goodin and Thomas Schillemans (eds), *The Oxford Handbook of Public Accountability* (Oxford University Press 2014).

transparency regulation and explore how it interacts with the domains of intellectual property law, data protection law, as well as how it is represented in the recently proposed draft Digital Services Act and AI Act. We then present the questions that regulators will methodically and purposefully need to tackle, in order to shape transparency to effectively yet proportionately safeguard individual autonomy.

## i. State of the art concerning transparency obligations

### *a. Intellectual property law*

Transparency of RS can both be facilitated and hindered by intellectual property (IP) law. Patents grant privileged rights to innovators in exchange for transparency that fosters scientific progress by sharing valuable and breakthrough insights. Mechanisms used in recommender and ranking systems may be patented[172] which could be a building block of their transparency. However, there are barriers to relying on patents for transparency. Firstly, the patentability of AI and ML is a matter of ongoing debate. The European Patent Convention excludes 'programs for computers' from being patentable inventions[173] and artificial intelligence and machine learning may be considered too abstract in nature to be patentable.[174] However, specific models that deliver a technical effect, such as targeting content to individuals in a particular manner, may be patentable if sufficiently innovative.[175] Secondly, even if patentable, the transparency provided by patents is targeted at experts, not average users. Disclosure need only be 'sufficiently clear and complete for it to be carried out by a person skilled in the art'.[176] This suggests extending software patentability may be part of the solution, but the transparency it provides, while valuable in advancing science, does not, as it stands, facilitate transparency to non-experts and lay persons. Finally, even where patents could deliver some form of transparency, the exclusive rights conferred

[172] Examples of patented recommendation and ranking algorithms include Google's PageRank, Facebook's newsfeed, and Amazon's multiple recommender system. Lawrence Page, 'Method for Node Ranking in a Linked Database'; Mark Zuckerberg and others, 'Communicating a Newsfeed of Media Content Based on a Member's Interactions in a Social Network Environment'; Jennifer A Jacobi, Eric A Benson and Gregory D Linden, 'Personalized Recommendations of Items Represented within a Database.'

[173] Convention on the Grant of European Patents of October 5, 1973 as revised by the Act revising Article 63 EPC of 17 December 1991 and the Act revising the EPC of November 29, 2000 (European Patent Convention) (adopted 5 October 1973) OJ EPO 2001, Special edition No. 4, 55, art 52(2)(c).

[174] European Patent Office, 'Guidelines for Examination in the European Patent Office' (March 2021) <https://www.epo.org/law-practice/legal-texts/html/guidelines/e/index.htm> accessed 31 March 2021 Part G, Chapter II, Point 3.3.1.

[175] ibid Part G, Chapter II, Point 3.3.

[176] European Patent Convention, art 83.

on innovators are likely to hinder competition and, thus, limit consumer choice and, by extension, the ability of individuals to exercise autonomy when choosing a recommender system.

Where RS are not patentable, innovators may look to other forms of protection. Copyrighting of the AI code can offer limited protection, since it does not extend to the principles and mechanics underlying the software, but rather just to the code as such.[177] Similar to how copyright protects written works of art, the words of Shakespeare's Romeo and Juliet as they are written are protected from being copied, however the story of two tragic lovers from rival families can be retold with different words and small changes to the story. In this same manner, copyright law cannot sufficiently protect the 'ideas' of how particular AI technologies operate, but rather simply protect their code, word-for-word. Handling AI models as trade secrets is often the choice of innovators. Trade secrets protect any information that is commercially valuable, including a method of production or an algorithm formula,[178] for as long as reasonable steps to keep it secret are maintained.[179] Perhaps the most notorious example of the conflict between transparency and trade secrets came with the US case of *Loomis v Wisconsin*, where a criminal defendant was denied access to a risk scoring algorithm used to inform the judge's decision in the case.[180] That case illustrates the tension between human rights and trade secrets. Such lack of transparency could place algorithms beyond the reach of legal assessments.[181] IP law, thus, can present a challenging field of law to navigate when discussing RS transparency to support individual autonomy. It either does not facilitate transparency for end users and non-experts or it hinders transparency overall.

### *b. Data protection law*

Transparency is also a key principle of personal data processing according to art 5(1)(a) of the GDPR - the principle of lawfulness, fairness, and transparency and, in that capacity, could help individuals understand more about what personal data of theirs is processed, how, and what rights they have

---

[177] Recital 11, Art 1(2) Directive 2009/24/EC on the legal protection of computer programs (Software Directive) [2009] OJ L 111, 16.

[178] Brian T Yeh, 'Protection of Trade Secrets: Overview of Current Law and Legislation' (*Congressional Research Service* 2016) <https://fas.org/sgp/crs/secrecy/R43714.pdf> accessed 6 March 2021.

[179] Art 2(1)(c) Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L157, 1.

[180] *Loomis v Wisconsin,* 881 NW2d 749 (Wis. 2016), cert. denied, 137 S Ct 2290 (2017)

[181] Woodrow Barfield and Ugo Pagallo, *Advanced Introduction to Law and Artificial Intelligence* (Edward Elgar Publishing 2020) 171 et seq.

in that regard. Recital 60 of the GDPR clarifies the obligation of data controllers to provide individuals with information that would be 'necessary to ensure fair and transparent processing.' This is further detailed in the regulation with a number of proactive transparency and disclosure requirements, as well as with rights on individuals to demand information (the right to access). All information regarding 'risks, rules, safeguards and rights' related to personal data processing and how to exercise rights should be clearly communicated with individuals.[182] Individuals should also be made aware of the existence and consequences of profiling,[183] particularly relevant to the manner of operation of RS. This information could, in theory, enhance individual awareness of external influences, however users often do not read the privacy policy documents where this information is recorded[184] and when they do, they might be confronted with vague or complicated text.

The GDPR could also provide individuals with glimpses into the process behind the creation and operation of the RS. The GDPR allows individuals to know how and by whom their personal data is handled and managed,[185] and also the purpose of the processing,[186] and they could receive copies of the personal data controllers hold about them.[187] This might help answer who, and for what reason, is processing personal data or profiling individuals, thus seeking to influence them. But there are limitations. Firstly, the specified purpose of processing might not reveal the specific goal of RS used. If personal data is collected to help improve a service, it is not clear what a RS would optimise for in order to improve such a service. Secondly, these rights would not allow individuals to know which of all of their personal data that is held by a controller are actually used or influential for the performance of RS. This limits the insight into the RS's logic that users could gain through data protection rights.

Finally, the GDPR could also limit what information about RS could be provided to individuals. It protects personal data from unauthorised disclosures.[188] Training data is of vital importance to the performance of ML algorithms, often used in RS. For that reason, it is important to consider whether,

---

[182] GDPR, arts 13(2), 14(2); Recital 39.
[183] GDPR, recital 60.
[184] Jonathan A Obar and Anne Oeldorf-Hirsch, 'The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services' (2020) 23 Information, Communication & Society 128; Nili Steinfeld, '"I Agree to the Terms and Conditions": (How) Do Users Read Privacy Policies Online? An Eye-Tracking Experiment' (2016) 55 Computers in Human Behavior 992.
[185] GDPR, arts 13(1)(a), 14(1)(a); Recital 39.
[186] GDPR, arts 13(1)(c), 14(1)(c); Recital 39.
[187] GDPR, art 15(3).
[188] GDPR, art 5(f).

and to what extent, training data should form part of relevant transparency obligations, with due regard to the data protection rights of individuals whose data may form part of such training data. Thus, the European data protection law, as it currently stands, also leaves potential gaps in terms of the use of transparency to support individual autonomy in a RS context.

### *c. Digital Services Act*

The draft DSA also promotes transparency, particularly in RS. It requires *very large* online platforms to disclose in a clear and accessible manner 'the main parameters used in their recommender systems.'[189] Transparency and usability, where there are options for users to modify or adjust the parameters of the RS, are also highlighted.[190] Here, transparency is used to empower users not only to understand the logic of the RS, but also to ensure that they can shape it. The DSA also makes strides with regard to targeted advertising transparency. Online advertisements are required to be clearly marked as such, notably including information on the identity of the natural or legal persons behind them, as well as 'meaningful information about the main parameters used to determine the recipient' of the advertisement.[191] While consumer protection law already mandates advertisements to be clearly marked,[192] this obligation would allow insights into the purpose and manner in which advertising seeks to target and influence individuals. Very large platforms would also have to publish aggregate data about advertising, including who ordered the advertisement, the intended recipients, as well as the number of recipients.[193] This might facilitate public accountability and research regarding advertising practices.

The draft DSA also proposes transparency obligations that shed light on the manner in which online platforms operate. A general obligation to be transparent about content moderation and handling of illegal content is discussed.[194] In addition, reporting duties for very large platforms are proposed,

---

[189] DSA, art 29(1).
[190] See s III.A. above and DSA, art 29.
[191] DSA, art 24.
[192] Arts 2(b), 3 Directive 2006/114/EC of the European Parliament and of the Council of 12 December 2006 concerning misleading and comparative advertising (codified version) 2006 [32006L0114] 21. ''Native Advertising: A Guide for Businesses | Federal Trade Commission' (*Federal Trade Commission*, December 2015) <https://www.ftc.gov/tips-advice/business-center/guidance/native-advertising-guide-businesses> accessed 6 March 2021; 'How to Comply with EU Rules Applicable to Online Native Advertising' (*IAB Europe,* 2016) <https://iabeurope.eu/wp-content/uploads/2019/08/IAB-Europe-Online-Native-Advertising-Guidance.pdf> accessed 6 March 2021.
[193] DSA, art 30.
[194] DSA, art 13.

covering *inter alia* their assessments of systemic risks arising out of their RS and targeted advertising systems[195] and proportionate and effective risk mitigation steps they have taken, including adapting their RS.[196] The performance of this and other obligations is subject to independent auditing[197] and reporting, also subject to disclosure.[198] The information that has been revealed to the public may be redacted to protect commercially confidential information or the privacy of other users, but would still be accessible to EU authorities.[199] Such disclosures could allow some transparency to users about the RS used and the way in which they operate, although it might be redacted or technical. However, public oversight through EU authorities would be ensured. This is a clear move towards embedding transparency into the operation of very large platforms with a prominent role for transparency of RS. However, even here, the types of transparency presented and their intended audiences are not all intended to safeguard the autonomy of individual users. Rather, the transparency obligations are framed as a way to ensure public oversight over the operation of highly impactful platforms.

### *d. The AI Act*

The recently proposed EU AI Act also lays down a number of requirements and obligations regarding the transparency and oversight of AI systems. Most of these obligations, however, only apply to systems which are classified as 'high risk.'[200] If RS meet this standard, then a range of transparency requirements would apply to them, including a documented and maintained risk management system,[201] design that is "sufficiently transparent to enable users to interpret the system's output and use it appropriately"[202] and that allows for human oversight[203] and technical documentation demonstrating compliance with high-risk AI requirements.[204] These requirements include information about the overall process of the system's creation, maintenance and oversight, relevant metrics, risks, and the system's design specifications, general logic, and "the key design choices including the rationale and assumptions made."[205] Thus, in the AI Act there are transparency obligations

[195] DSA, arts 26, 33.
[196] DSA, arts 26, 27(1)(a), 33.
[197] DSA, arts 28(3), 33.
[198] DSA, arts 28(4), 33.
[199] DSA, art 33(3).
[200] AI Act, annex III.
[201] AI Act, art 9.
[202] AI Act, art 13(1).
[203] AI Act, art 14.
[204] AI Act, art 9.
[205] AI Act, annexes IV.2.b and IV.

that cover both– the process of the system's creation, as well as its internal logic, architecture, performance and even extend to aspects of human-computer interaction (HCI). A publicly accessible EU database of high-risk AI systems is also envisioned, albeit containing less detailed information.[206] The draft AI Act also provides for some transparency obligations for AI systems that are not considered to be high-risk, however, these are more modest. For example, where natural persons interact with AI systems, they are informed of their AI nature.[207] This makes some, but not fully sufficient, progress towards ensuring individuals have enough information about RS to be aware of and fully understand the way RS influence them. Even if RS are considered "high-risk," not all of the information maintained about them is intended to be accessible to end-users or the public. Some of it is reserved for enabling oversight by public authorities, subject to appropriate confidentiality safeguards.[208] Nevertheless, a strong link to individual autonomy is the requirement to design "high-risk" AI systems in a manner that ensures humans can understand and use their outputs, thus putting users in an empowered position. There is, however, scope to further consider and develop transparency requirements for safeguarding autonomy of RS users.

### ii. Towards more meaningful transparency

Transparency obligations as well as limitations to transparency exist in a piecemeal manner across multiple legal frameworks. However, a coherent and purposeful approach would be necessary, using individual autonomy as the guiding "North Star" and goal of transparency. At the same time, regulating transparency also has to take into account competing interests, e.g. IP law, that may justify limited disclosure of information. This will require defining transparency and its relevant dimensions – scope of disclosure, obligations and rights, proactive or demand-driven disclosure, and intended recipients – in a purposeful manner that allows users to autonomously make their own informed decisions and also enables them to hold those that seek to influence them accountable. This could be done by (1) mediating the type and *content* of transparency obligations – both in terms of what is disclosed, as well as how it is disclosed, or (2) by moderating the *recipients* of information. These two aspects are interlinked, as information disclosed to a particular recipient should be understandable and usable by its intended recipient. Below, we highlight some of the challenges future regulation should account for.

---

[206] AI Act, art 60.
[207] AI Act, art 52(1).
[208] AI Act, arts 64(6), 70.

### *a. Defining the scope of transparency purposefully*

A first step for regulators would be to define what transparency should cover, with a view towards achieving specific objectives or goals. The content of disclosures may depend on the individual case – on its context and impact of the disclosure, as well as on the intended recipients of the information and their goals.[209] Having a clear view of what functionalities transparency should fulfil will help ensure that it is balanced and proportionate *vis-a-vis* competing interests. Regulators can then choose from a range of transparency options. A fundamental question is whether the goal of transparency is to inform end users and, thus, facilitate informed decision-making, or is it to facilitate human oversight of RS with the goal of indirectly protecting individual autonomy? Ideally, a complementary approach should be taken, taking advantage of the strengths of both approaches. For example, transparency can be achieved by providing information that does not infringe on trade secrets. Users can receive explanations about the RS's operation or a specific recommendation that allow them to understand the context and consequences of their actions, but that are not technical to the extent of breaching trade secrets.[210] Alternatively, if technical disclosures are necessary for an assessment of the RS, this may be done by limiting disclosure to authorised and independent organisations, similar to what is already practiced where public authorities examine commercially sensitive data, such as in IP litigation. There are mechanisms to allow the disclosure of sensitive information sufficiently to enable human control and oversight. Ideally, transparency regulation will seek to combine the strengths and complementarity of both approaches.

Once there is a clear goal for transparency to fulfil, regulators would need to narrow down the precise definition and scope of transparency that would allow them to achieve it. When it comes to a particular RS, we can differentiate between disclosing information about (1) the process of the creation of the RS - *process transparency* and (2) the results of the process - the system, its data and logic, performance, and results - *outcome transparency*.[211] Then, regulators need to consider what is *knowable* about algorithmic systems to identify the scope of desirable disclosure. What we can know about an

---

[209] Alan FT Winfield and Marina Jirotka, 'Ethical Governance Is Essential to Building Trust in Robotics and Artificial Intelligence Systems' (2018) 376 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20180085.

[210] Céline Castets-Renard, 'The Intersection Between AI and IP: Conflict or Complementarity?' (2020) 51 IIC - International Review of Intellectual Property and Competition Law 141.

[211] David Leslie, 'Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector' (*Zenodo* 2019) <https://zenodo.org/record/3240529> accessed 27 April 2020.

algorithm includes information about (1) human involvement and decisions made in the creation and implementation of a system, assumptions, goals, intents; (2) about the type, features, qualities, provenance and legal terms for the use of the data, as well as its management; and (3) about the model itself – its type, performance metrics, metadata (date, version), thresholds, assumptions, rules it includes, along with influential variables and weighting if known.[212] Regulators need to consider which of this information they would like disclosed to whom and in what shape in order to safeguard autonomy. As we saw above, the current and proposed legal framework provides for a tapestry of transparency and disclosure obligations.

Aspects of both outcome and process transparency are necessary to safeguard autonomy. Outcome transparency is key to support individual agency by highlighting how individual autonomy may be impacted. It could incorporate information about the model and how and why it operates. The Consultative Committee on the Council of Europe's Convention Hundred and Eight suggests that in order to enable public scrutiny, a reasonable solution could be disclosures of the logics of an AI algorithm in general, covering its overall operation, the type of expected input and output data, the variables and weights used by the algorithm, as well as details about its architecture.[213] Moreover, previous work on transparency in the context of nudging highlights the need for being transparent that a particular technique of nudging is used to achieve a particular goal,[214] as well as highlighting specific instances of nudging, making them identifiable to nudgees.[215] This is in line with the requirement in the EU's Guidelines for Trustworthy AI that AI systems be clearly identified as such to end-users along with information on the system's capabilities, limitations, and purpose[216] and can be encompassed within outcome transparency.

---

[212] Nicholas Diakopoulos, 'Transparency' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (Oxford University Press 2020) 201 et seq .

[213] Alessandro Mantelero, 'Artificial Intelligence and Data Protection: Challenges and Possibilities' (Consultative Committee of the Convention for the protection of individuals with regard to automating processing of personal data (Convention 108)) T-PD(2018)09Rev <https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6> accessed 6 March 2021.

[214] Cass R Sunstein, 'The Ethics of Nudging' (2015) 32 Yale Journal on Regulation 413; Luc Bovens, 'The Ethics of Nudge' in Till Grüne-Yanoff and Sven Ove Hansson (eds), *Preference Change: Approaches from Philosophy, Economics and Psychology* (Springer Netherlands 2009).

[215] Lembcke and others (n 16) 11.

[216] High-Level Expert Group on Artificial Intelligence (n 165) 18; High-Level Expert Group on Artificial Intelligence, 'Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment' (2020) Text 14 <https://ec.europa.eu/digital-single-market/en/news/

Instead, process transparency is vital to ensure accountability of those designing and creating RS and justifiability of the design choices made, for example what a RS is optimising for and which user data it considers influential. It could cover information about the human involvement in the RS's creation, as well as the data used and decisions made to tailor and optimise the model. Transparency of the data used is highlighted in a number of policies. In the EU's Guidelines for Trustworthy AI transparency should cover data traceability and provenance.[217] The Council of Europe, in its Report on AI, similarly highlights that transparency of data used to train and operate an algorithm.[218] Where information about the logic of operation of a RS may be unknowable due to its complexity (see point 2.c. below), process transparency could offer an important replacement mechanism of checks and balances. As the EU's Guidelines on Trustworthy AI suggest, where explanations of the way systems operate are not possible, other types of transparency should be prioritised.[219]

### *b. Understandable disclosure formats*

The goal of regulating for transparency should be to provide higher quality information rather than simply "more" information.[220] Information should be provided to its intended recipients in a useful manner and, following the example of the GDPR, should be easily accessible, understandable, concise, using "clear and plain language."[221] Regulation could standardise procedures and formats for disclosure,[222] including by considering mechanisms like standardised icons, certification schemes or seals.[223] Research from human-computer interaction (HCI) could help shed light on how information can be intuitively presented[224] or to help identify what explanations users and

assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> accessed 6 March 2021.

[217] High-Level Expert Group on Artificial Intelligence (n 165) 18.

[218] Mantelero (n 213) 11–12.

[219] High-Level Expert Group on Artificial Intelligence (n 216) 14–15.

[220] Rolf H Weber, 'Socio-Ethical Values and Legal Rules on Automated Platforms: The Quest for a Symbiotic Relationship' (2020) 36 Computer Law & Security Review 36:105380, 7.

[221] GDPR, art 12(1), Recitals 39, 58.

[222] Diakopoulos (n 212) 211.

[223] GDPR, art 12(7), Recital 100.

[224] Jaron Harambam and others, 'Designing for the Better by Taking Users into Account: A Qualitative Evaluation of User Control Mechanisms in (News) Recommender Systems', *Proceedings of the 13th ACM Conference on Recommender Systems* (*ACM* 2019) <https://dl.acm.org/doi/10.1145/3298689.3347014> accessed 6 March 2021; Chen He, Denis Parra and Katrien Verbert, 'Interactive Recommender Systems: A Survey of the State of the Art and Future Research Challenges and Opportunities' (2016) 56 Expert Systems with Applications 9; Dietmar Jannach, Sidra Naveed and Michael Jugovac, 'User Control in Recommender Systems: Overview and Interaction Challenges' in Derek Bridge and Heiner Stuckenschmidt (eds), *E-Commerce and Web Technologies: 17th International*

experts think are necessary, as well as how they can be provided.[225] In fact, there are already some existing tools that could help communicate outcome and process transparency. For outcome transparency, information about models can be provided through model cards, with an overview of model performance, its intended uses, limitations, and key architectural features.[226] Data transparency can be achieved by sharing 'definitions and meanings of variables in the data, as well as how they are measured'.[227] Documents like Datasheets or Dataset Nutrition Labels can play a role to record qualities of the data, as well as rationale for human manipulations.[228] Moreover, transparency of specific instances of recommendations could be achieved by highlighting them through the use of borders around elements or textual notifications. More research will be necessary to determine when digital elements on a page constitute a 'nudge' and how to best (visually) represent this to make individuals aware of it.[229] This may, however, be necessary especially for the DSA-proposed transparency and highlighting of targeted advertising. On the other hand, process transparency that provides insights into the creation of RS is also desirable. Information on human involvement can be collected progressively throughout the process of RS creation through end-to-end documentation intended to support accountability and auditability.[230] Relevant aspects for communication to individuals or authorities can then be extracted. This may require changes to internal work processes; however, this is nothing new. Legal acts, including the GDPR, often require both technical and organisational measures for compliance with their obligations.[231] It is important, however, that transparency regulation considers how to ensure disclosed information is useful and fit for purposes of safeguarding autonomy.

---

*Conference, EC-Web 2016, Porto, Portugal, September 5-8, 2016, Revised Selected Papers*, vol 278 (Springer International Publishing 2017).

[225] Malin Eiband and others, 'Bringing Transparency Design into Practice', *23rd International Conference on Intelligent User Interfaces* (*ACM* 2018) <https://dl.acm.org/doi/10.1145/3172944.3172961> accessed 6 March 2021.

[226] Margaret Mitchell and others, 'Model Cards for Model Reporting' (2019) Proceedings of the Conference on Fairness, Accountability, and Transparency 220.

[227] Diakopoulos (n 212) 203.

[228] Timnit Gebru and others, 'Datasheets for Datasets' [2020] arXiv:1803.09010 [cs] <http://arxiv.org/abs/1803.09010> accessed 6 March 2021; Sarah Holland and others, 'The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards' [2018] arXiv:1805.03677 [cs] <http://arxiv.org/abs/1805.03677> accessed 6 March 2021.

[229] Lembcke and others (n 16) 11.

[230] Inioluwa Deborah Raji and others, 'Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (*ACM* 2020) <https://dl.acm.org/doi/10.1145/3351095.3372873> accessed 6 March 2021.

[231] GDPR, Recital 78.

### *c. Explainability and oversight*

A final challenge that transparency regulation must tackle is the potential use of complex ML systems in ML. Where sophisticated ML algorithms are used, it may be impossible to know how and why systems operate the way they do[232] or how and why an individual has been classified a certain way and, therefore, receives a certain algorithmic output.[233] This problem, labelled 'black box' AI, addresses the necessity of explainable AI for trust and legal accountability.[234] Where AI is unexplainable, some types of transparency may be difficult to realise. However, there are two possible solutions. First, regulators may consider whether there is a need to limit the use of unexplainable and uninterpretable AI models.[235] Depending on the context, interpretability and transparency of AI models may be prioritised to ensure the legal compliance[236] of RS models used. For example, in the draft AI Act, high risk AI systems and their outputs have to be sufficiently interpretable to be used appropriately.[237] Some argue that interpretable models may perform just as well as 'black box' models,[238] with some initial supportive research in the area.[239] Second, where RS are uninterpretable, other information about the RS is still knowable. *Process transparency* is always possible. We might also disclose an algorithm's purpose or optimisation goal, design and basic

---

[232] Lembcke and others (n 16) 10.

[233] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' [2016] 3 Big Data & Society 205395171562251.

[234] Giulia Vilone and Luca Longo, 'Explainable Artificial Intelligence: A Systematic Review' [2020] arXiv:2006.00093 [cs] <http://arxiv.org/abs/2006.00093> accessed 6 March 2021; Finale Doshi-Velez and others, 'Accountability of AI Under the Law: The Role of Explanation' [2019] arXiv:1711.01134 [cs, stat] <http://arxiv.org/abs/1711.01134> accessed 6 March 2021; Finale Doshi-Velez and Been Kim, 'Towards A Rigorous Science of Interpretable Machine Learning' [2017] arXiv:1702.08608 [cs, stat] <http://arxiv.org/abs/1702.08608> accessed 6 March 2021.

[235] Lembcke and others (n 16); Burrell (n 233); Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (1st edn, Harvard University Press 2015).

[236] German AI Strategy specifically mentions the need for transparency in the way AI operates and produces outputs – the "criteria, objectives, logic" to assess compliance with legal requirements, including that of non-discrimination. German Federal Government, 'Artificial Intelligence Strategy' (2018) 16, 38 <https://www.ki-strategie-deutschland.de/home.html>.

[237] AI Act, art 13(1).

[238] Cynthia Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1 Nature Machine Intelligence 206; Cynthia Rudin and Joanna Radin, 'Why are we Using Black Box Models in AI When we don't need To? A Lesson From an Explainable AI Competition' (2019) 1 Harvard Data Science Review <https://hdsr.mitpress.mit.edu/pub/f9kuryi8> accessed 6 March 2021.

[239] Elaine Angelino and others, 'Learning Certifiably Optimal Rule Lists for Categorical Data' [2018] arXiv:1704.01701 [cs, stat] <http://arxiv.org/abs/1704.01701> accessed 8 April 2021.

functionalities,[240] such as the model's architecture and performance and could even cover the data processed on an individual basis. This information could be provided for oversight by sufficiently resourced public authorities. For example, "professional means, such as external auditors assessing the code (…) or (…) interdisciplinary partnerships" can be ways to ensure the ethical justifiability of uninterpretable algorithms used to shape individual choices.[241] The key is to provide information that enables meaningful human control that could then itself consider the impact of the system on autonomy.

In conclusion, transparency is a potentially powerful tool to safeguard autonomy; and has multiple dimensions that can be moulded by a regulator to achieve a desired purpose of enhancing autonomy. Technology-specific challenges of RS, such as uninterpretability, do not pose a barrier to all relevant transparency, nor do legal challenges, such as IP law or data protection law. By shaping the scope of transparency, its intended percipients, and usability, regulators could create a coherent framework to utilise the potential of algorithmic transparency for autonomy.

## IV. Conclusion

Recommender systems, by their very nature and intended use, affect individual autonomy and, boosted by profiling and micro targeting, are able to shape human thought and action. Ultimately, this affects individual and collective autonomy and self-determination, as well as human rights. The current regulatory framework, as it exists, leaves gaps in terms of ensuring accountability and oversight of the creation, use, operation, and impacts of RS. Exercising such power, however, cannot be permissible without appropriate checks and balances. In this paper, we mapped the current and recent European legislative trends with relevance to RS and their impact on autonomy to highlight how, through different angles and with different justifications, there is a clear indication that this is an issue very much on the policy agenda. We proposed a set of considerations and possibilities for the future development of a regulatory framework that can appropriately control the exercise of such power over user autonomy. We structured our analysis to address RS's design (Section III.A), the data they use (Section III.B), and the information about them which is presented to end users or qualified third parties (Section III.C). Key steps that can serve to safeguard or promote

---

[240] Cary Coglianese and David Lehr, 'Transparency and Algorithmic Governance' (2019) 2123 Faculty Scholarship at Penn Law 1.

[241] Lembcke and others (n 16) 10; Pasquale (n 235); Burrell (n 233).

individual autonomy are possible at each of these junctions in the creation and operation of RS.

A possible autonomy-by-design approach could empower the self-determined and directed use of RS by individuals, aligning RS with the general preferences of users. This can be enabled *inter alia* by architectures of user control, user shaping algorithms or choosing between algorithms. This idea is already foreshadowed in current legislative proposals and in projects in the IT-industry. Moreover, technical additions or modifications of RS may also diminish the manipulative impact that RS have on individual experiences, e.g., by including serendipity or randomisation techniques. A rights-based approach could also empower users to control the way their data is processed and their digital reflections and the GDPR can help mitigate some of the risks created by RS[242] through the data subjects' rights and principles and obligations for controllers it establishes. However, as seen throughout this article, there are still blind spots, for instance, in terms of effective application in practice or control of individuals over inferences made about them, meaning that individuals cannot fully control the processing of their data by RS, nor the impact RS have on them. Additional consideration of impact assessments, audits, greater transparency, freedom of choice, and even new rights may be necessary to effectively close this gap. Finally, transparency, through its role in empowering user choice, understanding, and accountability of RS creators and deployers, also has a vital role to play. Yet, it has to be regulated in a complex landscape of overlapping and conflicting interests and obligations that include IP law, data protection law, and features of complex AI technologies. Nevertheless, a push towards transparency is visible in the recent legislative initiatives in the EU. To help regulators think through this complex field going forward, we emphasise the need for purpose-driven regulation and build a taxonomy of the diversity of information, recipients, and forms of disclosure that regulators can consider when shaping their policy.

The regulatory options we have highlighted should not prejudice other complementary and vital efforts. One such effort is investing in digital literacy and education to enhance people's awareness and knowledge about artificial intelligence,[243] especially regarding the harms, benefits and effects of the most common applications that track, target and categorize individuals

[242] Access Now, 'Human Rights in the Age of Artificial Intelligence' (2018) 30 <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf> accessed 31 March 2021.

[243] Commission, 'Digital Education Action Plan 2021-2027: Resetting Education and Training for the Digital Age' (European Commission 2020) COM(2020) 624 final 4 <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0624&from=EN>.

(RS, for instance). It is vital to reduce information asymmetries, empower users and make them more prepared to deal with these technologies, ensuring their rights and informational self-determination. Another example is interdisciplinary training of data scientists that could improve the value-driven design of technology in practice. Currently, discrepancies between formal legal requirements and the real practice can reduce the GDPR to a formality.[244] It requires a more granular application of its rules,[245] in a "user-centric design".[246] Careful in-depth consideration throughout the process of technology design and implementation is necessary in order to ensure desirable social outcomes are achieved and relevant legal standards or goals are met. In order to safeguard individual autonomy from the continuous shaping and moulding exerted by RS online, regulators and policy-makers need to think holistically about the different dimensions of regulation, as well as the entirety of the RS— from their creation and design, through their deployment, until their final use and interaction with human beings.

Whether it is a law-by-design approach that seeks to shape RS design, a rights-based approach to empower users to control how they are perceived and profiled by RS, or a focus on processes and procedures to correct asymmetries in information and power between creators and users of RS through transparency measures, there are a range of tools available to policy makers. Like any regulation, turning these ideas into a regulatory framework would require balancing competing interests. Despite the potential challenges, there needs to be a clear stance about the priority of individual autonomy as a value that is worth pursuing and protecting. Autonomy and self-determination, both individual and collective, underpin fundamental values in our social and legal orders, including the rule of law, democracy, and human rights. As the digital increasingly shapes large parts of our lives, the protection of autonomy needs to be expanded and cover operation of innovations exerting 'soft power' over us like RS. Using the goal of individual autonomy as a North Star to aim for, policy-makers could shape a purposeful regulatory space that ensures truly human-centred technology. This is a young and dynamic field of research, however, and more is undoubtedly to come. New policy developments, such as the drafts of the Digital Services Act, the AI Act, and the e-Privacy Regulation, clearly highlight that there is political will to act and shape technology instead of simply allowing it to shape us.

---

[244] Giannopoulou (n 115) 4–6.

[245] ibid 6; Solon Barocas and Helen Nissenbaum, 'Big Data's End Run around Anonymity and Consent' in Julia Lane and others (eds), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (Cambridge University Press 2014) <http://ebooks.cambridge.org/ref/id/CBO9781107590205> accessed 6 March 2021.

[246] Giannopoulou (n 115) 9.

To make these attempts meaningful, it is important to focus on autonomy through different means.