IMT School for Advanced Studies, Lucca
Lucca, Italy

User Empowerment in the Digital Services Act: Redirecting Platforms' Influence through Transparency and Controllability of Recommender Systems

PhD Program in Cybersicurezza

Track in Human, Economic, and Legal Aspects in Cybersecurity

XXXVIII Cycle

By

Matteo Fabbri

2026

The dissertation of Matteo Fabbri is approved.

PhD Program Coordinator: Mirco Tribastone, IMT School for Advanced Studies Lucca

Advisor: Prof. Andrea Simoncini, University of Florence

Co-Advisor: Prof. Ludovico Boratto, University of Cagliari

The dissertation of Matteo Fabbri has been reviewed by:

Hans-Wolfgang Micklitz, European University Institute

Nava Tintarev, Maastricht University

IMT School for Advanced Studies Lucca
2026

A mio nonno Floriano,
per la nave carica d'oro che continua il suo viaggio.

Contents

List of Figures

List of Tables

Acknowledgements

Developing an interdisciplinary PhD thesis based on an individually designed research project is a challenging yet rewarding task, often blurring the boundaries between personal and professional life. First, to those that have supported or been close to me as their (grand)child, relative, friend, partner - and in any other personal capacity - thank you: I guess you know who you are. A special mention is deserved also by those that, either personally or professionally, have made my path more difficult than expected: thanks to them, I have improved my ability to identify what and who is worth pursuing or avoiding in academic life and beyond.

From a scientific perspective, I would like to thank my supervisor Andrea Simoncini for the freedom and flexibility which was granted to my research efforts: as I defined my own project, I had the possibility of choosing its directions without constraints or fixed guardrails. My co-supervisor Ludovico Boratto has provided invaluable guidance and support during the last year of my PhD, which coincided with its most critical phase: after meeting by chance, our shared interest and multi-sided understanding of recommender systems led to a precious collaboration.

I formally acknowledge the co-authors of the published and submitted papers of mine whose text contributes to thesis chapters, as detailed in the Introduction: Ludovico Boratto, Michele Loi, Andrea Ferrario, Jingyi Jia, Pablo Jerez Arnau and Wolfgang Woerndl. In particular, the collaboration with the latter three has been essential for building the technical foundation of the empirical pillar of my dissertation.

Vita

June 17, 1998	Born, Ferrara, Italy
2020	Bachelor's Degree in Philosophy Final mark: 110/110 cum laude University of Bologna, Bologna, Italy
2022	Master's Degree in Sociology and Global Challenges Final mark: 110/110 cum laude University of Florence, Florence, Italy
2022	Diploma di Licenza in Social and Political Sciences Final mark: 100/100 cum laude Scuola Normale Superiore, Florence, Italy
2022	MSc in Social Science of the Internet University of Oxford, Oxford, United Kingdom
2022 - present	PhD Candidate in Cybersecurity IMT School for Advanced Studies Lucca, Lucca, Italy
2023	Intern in Legal AI BMW Group, Munich, Germany
2024	Scientific Trainee at the Joint Research Centre European Commission, Ispra, Italy
2025	Visiting Scholar at the Institute for Information Law University of Amsterdam, Amsterdam, Netherlands

Publications

1. Fabbri, M. (2023). Social influence for societal interest: a pro-ethical framework for improving human decision making through multi-stakeholder recommender systems. *AI & Society*, 38(2), 995-1002. https://doi.org/10.1007/s00146-022-01467-2
2. Fabbri, M. (2023). Self-determination through explanation: an ethical perspective on the implementation of the transparency requirements for recommender systems set by the Digital Services Act of the European Union. *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 653–661. https://doi.org/10.1145/3600211.3604717
3. Fabbri, M. (2023). How and to which extent will the provisions of the Digital Services Act of the European Union impact on the relationship between users and platforms as information providers?. *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 1002–1003. https://doi.org/10.1145/3600211.3604749
4. Blanchini, M., Cerreta, M., Di Monda, D., Fabbri, M., Raciti, M., Ahmad, H. S., & Costa, G. (2024). Supporting Criminal Investigations on the Blockchain: A Temporal Logic-based Approach. *ITASEC 2024: The Italian Conference on CyberSecurity*. https://ceur-ws.org/Vol-3731/paper33.pdf
5. Reviglio, U., & Fabbri, M. (2024). Navigating the Digital Services Act: Scenarios of transparency and user control in VLOPs' recommender systems. *NORMalize 2024: The Second Workshop on the Normative Design and Evaluation of Recommender Systems, October 18, 2024*, co-located with the *ACM Conference on Recommender Systems 2024 (RecSys '24)*, Bari, Italy. https://ceur-ws.org/Vol-3898/short2.pdf
6. Fabbri, M. (2025). The Role of Requests for Information in Governing Digital Platforms Under the Digital Services Act: The Case of X. *Journalism and Media*, 6(1), 41. https://doi.org/10.3390/journalmedia6010041
7. Loi, M., Fabbri, M., & Ferrario, A. (2025). Pluralism isn't Just Methodological, it's Political: A Response to Wörsdörfer. *Philosophy & Technology*, 38(2), 1-3. https://doi.org/10.1007/s13347-025-00915-3
8. Loi, M., Fabbri, M., & Ferrario, A. (2025). Regulating the undefined: Addressing systemic risks in the digital services act (with an appendix on the AI act). *Philosophy & Technology*, 38(2), 1-29. https://doi.org/10.1007/s13347-025-00903-7

9. Fabbri, M. (2025). A classification framework for algorithmic recommendations: system, agent, patient. *Proceedings of the Fourth International Conference on Hybrid Human-Artificial Intelligence*. https://doi.org/10.3233/FAIA250661

10. Fabbri, M., & Boratto, L. (2025). Auditing Recommender Systems for User Empowerment in Very Large Online Platforms under the Digital Services Act. *Proceedings of the Nineteenth ACM Conference on Recommender Systems (RecSys '25)*, September 22–26, 2025, Prague, Czech Republic. https://doi.org/10.1145/3705328.3748074

11. Fabbri, M. Commento all'articolo 50 AI Act. In: Mantelero, A., Resta, G., & Riccio, G. M. (a cura di) (2025). *Intelligenza artificiale: Commentario. Regolamento (UE) 2024/1689 – AI Act, linee guida Commissione europea, pratiche vietate, Disegno di legge sull'intelligenza artificiale*. Wolters Kluwer Italia. ISBN 978-8821786013

12. Fabbri, M., Jia, J., Jerez Arnau, P., & Woerndl, W. Feeding the (short-video) feed: a design proposal for user control of social media recommender systems under the Digital Services Act. Under review for *International Journal of Human-Computer Studies*

Presentations (selected)

This section does not include the contributions to venues with published proceedings already mentioned among the publications.

1. "An ethical perspective on the new transparency requirements for recommender systems set by the Digital Services Act". *Conference on AI, Human Values and Meaningful Human Control*, Centre for Science and Thought, University of Bonn, Bonn (2023).
2. "Reshaping digital nudging after the Digital Services Act: from recommender systems as trust enablers to user-informed recommendations". *ITASEC 2024: The Italian Conference on CyberSecurity*, Salerno (2024).
3. Kick-off intervention (invited). *DSA Observatory Joint Workshop on Risk Management and Researchers' Data Access under the DSA*, Institute for Information Law, University of Amsterdam, Amsterdam (2025).
4. "Meaningful personalization as a guiding principle for the user control of recommender systems". *Empowering People in Online Spaces: Democracy and Well-being in Digital Societies (7th Weizenbaum Conference)*, Weizenbaum Institute, Berlin (2025).
5. "Who controls recommender systems? Algorithmic transparency and users' rights in (very large) online platforms after the Digital Services Act" (invited). *Workshop "How Computing is Changing the World: Exploring Synergies and Challenges of AI and Quantum Technologies for Society"*, University of Amsterdam, Amsterdam (2025).

Abstract

This thesis examines how the Digital Services Act (DSA)'s provisions on the transparency and controllability of recommender systems (RSs) empower users to redirect the influence that platforms exert on them through three interconnected disciplinary pillars. The philosophical pillar connects the user empowerment principle grounding the DSA with an ontology of influence of RSs: this conceptualization is used to argue in favour of user control as a design paradigm on which a substantive application of the DSA provisions should be based to address the harms generated by RSs. Secondly, the legal pillar outlines the main EU legal sources concerning RSs and discusses their practical implementation by examining the respective stages of platform compliance and regulatory enforcement. Subsequently, legal analysis and design are bridged to propose principles for meaningful personalization through RSs and a transparent and pluralistic enforcement of the DSA systemic risk framework. Thirdly, the empirical pillar highlights how users use control options for RSs and perceive the impact of these controls on their experience through studies on a realistic purpose-built short-video platform. The original design of the control features for short-video recommendations of the platform operationalizes the principles of proportionality and granularity introduced in the legal pillar, thereby representing an initial blueprint for the standardization of the DSA requirements on the transparency and controllability of RSs. All together, the three pillars provide a multi-layered *core sample* of the complex and rapidly evolving field of RS regulation, with each interrelated perspective illuminating a distinct dimension of its development.

Chapter 1

Introduction

This chapter highlights the topic and argumentative line of the thesis. It analyses, as an introductory case study, the first judicial application of the provisions of the Digital Services Act on the transparency and controllability of recommender systems. Building on the principle of user empowerment emerging from the implementation of these provisions, it outlines the research question directing the thesis and the respective contributions of its philosophical, legal and empirical pillars.

1.1 Context

On October 2nd 2025, the District Court of Amsterdam issued a ruling for the Bits of Freedom vs Meta Ireland case (Bits of Freedom, 2025), concerning the lack of direct and easy accessibility and persistence of users' selection of the non-profiling option for Facebook and Instagram's recommender systems (RSs) under the Digital Services Act (DSA) (European Parliament and Council, 2022). The Dutch NGO Bits of Freedom (BoF) sued Meta as, in its view, it would infringe the requirements of art. 25, 27 and 38 DSA based on the following arguments (District Court of Amsterdam, 2025):

1. "The place to opt for a non-profiled recommendation system is difficult to find and therefore not directly and easily accessible on the platforms.

2. After a user opens the platforms (the apps and/or websites), a profiled recommendation system is automatically applied to the home page, the comments section, and the reels section, even if the user selected a chronological recommendation system before the app/website was last closed. This 'switch back' to a profiled recommendation system also happens automatically during a session when a user navigates between different sections of the platform (the home page, the comments section, or the reels section).

3. It is impossible to set a 'persistent' choice in the apps or on the websites. That is to say, a choice to set a non-profiled recommendation system that does not 'switch back' to a profiled recommendation system when the user navigates between different sections of the platform or closes and reopens the platform's app or website.

4. If the user chooses a chronological recommendation system for the platforms' home page, the user ends up in a significantly different interface, in which essential functionalities are not accessible."

The court, the application of whose ruling is limited to Meta Ireland, the subsidiary of Meta Inc. in the EU, found the service provider not compliant with art. 27(3) and 38 (respectively requiring platforms to provide an accessible functionality for users to control RSs and at least one not option for recommendations not based on profiling), agreeing with BoF's arguments, because "the Instagram home page on the Android app, the reels section of Instagram (on both apps and the website), and the home page and reels section of Facebook (on both apps and the website) do not meet the requirements" of direct and easy accessibility and persistence of users' selection of the option for the non-profiled RS. This judgement brings about an important clarification regarding the interpretation of art. 27(3). In fact, the regulatory text does not explicitly state that the options for RSs selected by users should be persistent, i.e. that they should not change after users' selection, even when closing the application or switching to a different section of the platform. However, the court followed BoF's argument according to which "the autonomy and freedom of choice of users in choosing a recommendation system is fundamental to the exercise of freedom of information gathering, the fundamental right that in turn is fundamental to a democratic process such

as elections". As the selection of a non-profiled RS "requires an active action on the part of the user", it is not plausible to claim that "navigating between the different sections of the platform [...] constitutes a deliberate choice to undo that selection and opt for a profiled recommendation system again". Therefore, requiring users "to make a choice again and again (each time the apps and/or website of the platforms are opened) [...] is not compatible with a textual interpretation of Article 27(3) of the DSA" and, as "it leads to choice fatigue and an infringement of users' autonomy", it is also "contrary to Article 25 of the DSA and at odds with the purpose of Articles 27(3) and 38 of the DSA".

Given these motivations, the judges ordered Meta Ireland to "to respect and persistently apply user preferences for recommendation systems, meaning that a choice made by Dutch users for a non-profiled recommendation system will be retained, even if the user navigates to other sections within the platform, and even if the user closes and then reopens the apps or website". Moreover, the provider is also required to "make the preferred option for a non-profiled recommendation system directly and easily accessible on (i) the home page of the Android app, (ii) the reels section of Instagram (all apps and websites) and (iii) the home page and reels section of Facebook (all apps and websites)". The court also specified that the blueprint for how an easily accessible functionality to disable profiling-based recommendations should look like on the platforms' interface is represented by "the comments section of Facebook and Instagram, where the user can directly click/press on 'For you' or 'Latest' respectively". While Meta announced its intention to appeal the decision, the order to change its interface remains binding.

It is worth noting that an official interpretation of what the requirements of art. 27(3) entail, along with specific design suggestions on how to implement them, has not come from the European Commission yet, which is the main enforcer of the DSA against Very Large Online Platforms and Search Engines (VLOPSEs) such as Facebook and Instagram. The fact that this first-of-its-kind legal decision on RSs design and user control emerges from private enforcement of the DSA instead of the Commission and the national Digital Services Coordinators (DSCs) has

been noted by Meta, which stated that "the issue is a matter for the European Commission and regulators at the European levels and not for courts in individual countries" (Reuters, 2025). In fact, private enforcement, together with out-of-court dispute settlement, is one of the mechanisms through which the DSA can be applied outside of the enforcement actions led by the Commission and subject to the providers' appeal at the Court of Justice of the European Union (CJEU). While not legally sound, Meta's argument leads to question why the Commission's enforcement and secondary regulation have not substantially touched upon this matter as of now.

An initial answer might refer to the Commission's cautious approach to enforcing certain aspects of the DSA in the current geopolitical context, especially considering the "ongoing pressure from the Trump administration to challenge the implementation of European digital regulations" (Jahangir, 2025). However, given the large fines already imposed on US tech companies such as Apple and Meta under the Digital Markets Act (Moens, 2025), this answer does not seem sufficient to clarify the set of reasons for a lack of regulatory guidance on the implementation of the DSA obligations on RSs, which remain vague to a large extent. Another plausible explanation focuses on the complexity of these requirements, which require a technical grounding to be implemented and a solid knowledge base to be understood, which the regulator is currently lacking. Indeed, the DSA enforcement team of DG Connect opened a call for tender in August 2025 "to develop and deliver comprehensive user models that capture the complex interactions between users and recommender systems on online platforms, specifically VLOPs and VLOSEs" (European Commission, 2025a). Therefore, the regulator may plan to set guidelines on how "to give users more control and understanding of how information is presented to them on online platforms, particularly in relation to personalized recommendations" based on the scientific evidence emerging from the tenderer's research.

In the case at hand, the results of the summary proceedings initiated by BoF apply only to Dutch users, as the decision was issued by a national court. However, this case constitutes a precedent to which other

national courts, the CJEU or the Commission might refer to for their decisions, especially because it provides the first legally binding interpretation of art. 27(3) and an instance of application of the concept of dark pattern under art. 25 to an interface feature of two VLOPs. BoF was able "to obtain a decision within one and a half weeks" (Jahangir, 2025) due to the impact that RSs have on electoral processes within the systemic risk framework of art. 34, as the national elections in the Netherlands are scheduled for 29th October 2025. This urgency also justifies the fact that Meta has to comply with the court's decision in two weeks starting from October 2nd, the date of issuance of the judgement.

To challenge the extent and timing of the decision, Meta Ireland used two arguments that are seemingly inconsistent with each other. First, it argued that "any measures imposed should only apply until the elections, because after that the urgent interest will no longer apply" (District Court of Amsterdam, 2025). Second, it stated that "it is impossible for it to implement these orders within two weeks, as this would entail radical changes to its platforms, the implementation of which would take at least six to twelve months". While the first argument does not stand because providers should be compliant with the DSA at any time, and Meta was found non-compliant, the first and second arguments considered together clearly underline the provider's intention not to comply at all. In fact, if Meta claims that the measures should apply only until the Dutch elections but, at the same time, it argues that to implement the required technical changes it needs at least six months, it is de facto arguing that the court's order cannot be respected. Such an argumentative approach highlights the real issue at stake: "Meta is not providing users the autonomy required under the law" by making user control ineffective by design.

On 6th January 2026 Meta implemented the required interface changes, with a delay of six days with respect to the extended deadline set on 31st December 2025; however, "for approximately 4% of Facebook and/or Instagram users, the changes will not yet be visible until mid-March 2026, due to ongoing identification of glitches, bugs, or other technical issues resulting from the implemented changes" (Amsterdam Court of

Appeal, 2026). The company appealed the first court's decision in October 2025, but eventually decided to contest only procedural aspects - particularly the "urgent nature of the lawsuit" - after withdrawing all the other objections one day before the hearing (Bits of Freedom, 2026). The appeal ruling of 10th March 2026 confirmed "the contested ruling in all other respects" apart from the amount of the maximum sanction for non-compliance, which the judges doubled to 10 million euros (Amsterdam Court of Appeal, 2026). This means that the new chronological feed will remain in place for Instagram and Facebook users in the Netherlands.

1.2 Contributions

The District Court of Amsterdam (2025) highlights that "autonomy, freedom of choice and control over the way in which information is presented" are the principles grounding the EU legal framework on RSs. Indeed, as the main tool through which citizens access information online, RSs drive opinion formation processes (Gandini, Keeling, and Reviglio, 2025) while also contributing to the economic and political purposes of the tech companies that own them (Bodo, 2025). While users have been traditionally regarded as the passive stakeholder of the recommendation process, i.e. those consuming the recommended content, the DSA aims to empower them to "influence how information is presented to them" (recital 70): the role of transparency and user control for RSs is therefore to enable users to influence the socio-technical infrastructure that has been influencing them since the rise of social media and other online platforms. However, reverting the role of users from a passive to an active one is not straightforward: in fact, not only the mechanisms through which RSs influence users are to, a large extent, still obscure due to their black-box and proprietary nature, but it also needs to be understood whether users would actually make use of the increasing control that the DSA, at least in theory, grants them. Therefore, the research question guiding my thesis is: **How can the DSA provisions on the transparency and controllability of RSs empower users to redirect the influence that platforms exert on them, in order to self-determine**

their online experience?

My thesis addresses this gap through an interdisciplinary approach based on three complementary pillars, each guided by a sub-question reflecting a specific dimension of the issue at hand:

- **Philosophical pillar**: How can an ontological framing of RSs address their distinctive influence on user agency and provide a foundation for the EU regulatory approach to user empowerment?

- **Legal pillar**: How and to what extent does the implementation of the DSA provisions on RSs, from the side of platforms and regulators, enhance user empowerment through transparency and direct control?

- **Empirical pillar**: How do users engage in controlling RSs if given appropriate tools to do so?

I structured the thesis around these three disciplinary pillars because each of them offers a peculiar yet complementary perspective on the main research question. Firstly, the philosophical background provided in the next chapter is an original attempt to connect the user empowerment principle grounding the DSA with an ontology of influence of RSs: this conceptualization is used to argue in favour of user control as a design paradigm on which a substantive application of the DSA provisions should be based to address the harms generated by RSs. Secondly, the analysis of Chapters 3 to 5 not only outlines the main EU legal sources concerning RSs, but also discusses their practical implementation by examining the respective stages of platform compliance and regulatory enforcement. Chapters 4 and 5 bridge legal analysis and design by proposing principles for meaningful personalization through RSs and a transparent and pluralistic enforcement of the DSA systemic risk framework. Thirdly, the empirical studies presented in Chapters 6 and 7 highlight how users use control options for RSs in a realistic purpose-built short-video platform and how they perceive the impact of these controls on their experience. The original design of the control features for

short-video recommendations of the platform operationalizes the principles of proportionality and granularity introduced in Chapter 4, thereby representing an initial blueprint for the standardization of the DSA requirements on the transparency and controllability of RSs. All together, the three pillars provide a multi-layered *core sample* of the complex and rapidly evolving field of RS regulation, with each interrelated perspective illuminating a distinct dimension of its development.

In Chapter 2 [1] I build the philosophical pillar through an ontological categorization allowing to distinguish RSs from other types of algorithmic systems providing recommendations and I trace a link between each category and its application to relevant EU digital regulations. This categorization represents the conceptual foundation on which the influence of RSs on user agency is considered from the perspectives of nudging, persuasion and manipulation. Ethical implications regarding the harms of RSs are subsequently discussed to draw a link between with the DSA framework, in which these algorithms are considered enablers of systemic risks.

Chapters 3 to 5 shape the legal pillar. Chapter 3 [2] describes the EU

[1] Chapter 2 is based on the following published papers of mine:

- For section 2.2: Fabbri, Matteo (2025). "A Classification Framework for Algorithmic Recommendations: System, Agent, Patient". In: *Frontiers in Artificial Intelligence and Applications, Volume 408: HHAI 2025*. Amsterdam, The Netherlands: IOS Press, pp. 449–454 (Fabbri, 2025a).
- For section 2.3: Fabbri, Matteo (2023). "Self-determination through explanation: an ethical perspective on the implementation of the transparency requirements for recommender systems set by the Digital Services Act of the European Union". In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 653–661 (Fabbri, 2023b).

[2] Chapter 3 is based on the following published papers of mine:

- For sections from 3.1 to 3.5: Fabbri, Matteo (2023). "Self-determination through explanation: an ethical perspective on the implementation of the transparency requirements for recommender systems set by the Digital Services Act of the European Union". In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 653–661 (Fabbri, 2023b).
- For subsection 3.2.1: Fabbri, Matteo and Ludovico Boratto (2025). "Auditing Recommender Systems for User Empowerment in Very Large Online Platforms under the Digital Services Act". In: *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, pp. 51–61. (Fabbri and Boratto, 2025).

legal framework for the transparency and user control of RSs, reviewing relevant provisions of the DSA (and cognate regulation), the AI Act and the GDPR. Chapter 4 [3] evaluates the state-of-the-art of VLOPs' compliance with the DSA provisions on RSs through an analysis of the audit and systemic risk reports of Instagram, TikTok and YouTube. What emerges from the reports is that platforms fail to provide substantive interpretations of the legal requirements and align towards minimal compliance. Building on these results, I outline a perspective for meaningful personalization through algorithmic choice, balancing proportionality and granularity, and content curation, ensuring diversity and authoritativeness. Chapter 5 [4] outlines the unfolding of the DSA enforcement process through a case study of the phases of the Commission's investigation on X's compliance. The evidence emerging from this analysis motivates a critique of the opacity of the Commission's decisions leading to the initiation of proceedings. The risks and potential consequences of this approach are exemplified through a theoretical scenario based on shadow-banning through RSs. I then propose a constructive dual-track approach to the evaluation of systemic risks for both compliance assessment and researchers' data access.

- For subsection 3.2.4: Loi, Michele, Matteo Fabbri, and Andrea Ferrario (2025). "Regulating the Undefined: Addressing Systemic Risks in the Digital Services Act (with an Appendix on the AI Act)". In: *Philosophy & Technology* 38.2, p. 82 (Loi, Fabbri, and Ferrario, 2025).

[3] Chapter 4 is fully based the following published paper of mine:

- Fabbri, Matteo and Ludovico Boratto (2025). "Auditing Recommender Systems for User Empowerment in Very Large Online Platforms under the Digital Services Act". In: *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, pp. 51–61. (Fabbri and Boratto, 2025).

[4] Chapter 5 is based on the following published papers of mine:

- For section 5.2: Fabbri, Matteo (2025). "The Role of Requests for Information in Governing Digital Platforms Under the Digital Services Act: The Case of X". In: *Journalism and Media* 6.1, p. 41. (Fabbri, 2025b)
- For section 5.3: Loi, Michele, Matteo Fabbri, and Andrea Ferrario (2025). "Regulating the Undefined: Addressing Systemic Risks in the Digital Services Act (with an Appendix on the AI Act)". In: *Philosophy & Technology* 38.2, p. 82 (Loi, Fabbri, and Ferrario, 2025).

Chapters 6 and 7 are devoted to the empirical pillar. Chapter 6 [5] describes the design and implementation of a controllable and transparent RS integrated into an interactive interface based on YouTube Shorts, through which a user study was conducted to provide insights into how DSA-informed RS control features can enhance users' understanding and willingness to intervene on the recommendation process. Chapter 7 introduces an updated version of the interface on which user surveys and interviews were conducted to directly assess how users interact with control tools and reflect on how they can impact their online experience. The choice to focus on short video platforms, both for the analysis of the audit and systemic risk reports in Chapter 4, and for the user studies of Chapters 6 and 7, are due to the impact that this kind of content has on young users, including minors, whose protection under art. 28 DSA has led to the release of the first guidelines addressing RSs design, transparency and control.

The implications emerging from the three pillars are collected in Chapter 8, the conclusion, which traces a perspective on how the DSA is expected to impact on users' relationship with RSs in online platforms.

The options through which individuals can directly shape recommendations are not mere design elements, but rather socio-technical mechanisms that allow users to exercise their autonomy in choice architectures that deeply influence their onlife (Floridi, 2015b). In the current geopolitical context, RSs are infrastructures of influence considered strategic also for national security. In September 2025, the Trump administration pushed for a deal to sell TikTok's US operations to US investors, including Oracle, due to the concern that "the algorithm that fuels what users see on the app is vulnerable to manipulation by Chinese authorities, who can use it to shape content on the platform in a way that's difficult to detect" (Madhani, 2025). One of the crucial aspects of the negotiation with

[5] Chapter 6 is fully based on the following paper of mine, currently under the third round of review for the *International Journal of Human-Computer Studies*:

- Fabbri, Matteo, Jingyi Jia, Pablo Jerez Arnau and Wolfgang Woerndl. "Feeding the (short-video) feed: a design proposal for user control of social media recommender systems under the Digital Services Act".

the Chinese government involves the ownership of TikTok's RS: in fact, according to the acquisition plan outlined by Trump's executive order, "the algorithm will be retrained and monitored by the U.S. company's security partners, and operation of the algorithm will be under the control of the new joint venture" (Shepardson and Alper, 2025). Who controls RSs, on both a macro and micro level, is a matter that has wide-ranging implications for platform governance, AI ethics and digital regulation.

Chapter 2

Philosophical Background

This chapter provides a philosophical account of the influence of RSs. It first advances an ontological framework to distinguish RSs from other AI systems providing recommendations. Building on this categorization, it characterizes the relationship of influence between users and RSs as one of reliance of the former on the latter, which the user empowerment principle emerging from the DSA aims to counterbalance. Subsequently, it proposes an ethical perspective connecting user reliance on RSs to the harms enabled by them, which correspond systemic risks in the DSA formulation.

2.1 Introduction

In the contemporary digital age, RSs have become the dominant mediators of human attention in digital environments: from social media to music streaming, from e-commerce to news feeds, recommended content and products get the spotlight on interfaces and influence individuals' interests and priorities. From a technical perspective, RSs can be defined as algorithms aimed at estimating predictive ratings for items which a user has not seen yet (Adomavicius and Tuzhilin, 2005) in order to generate recommendations about content that may interest them. Therefore, RSs filter and prioritize information in ways that determine what users see, believe, and ultimately choose: this algorithmic media-

tion represents a digital infrastructure of epistemic dependence. While algorithmic recommendations, as the output of RSs, are aimed at improving user experience by reducing the information overload, they can give rise to ethical concerns related to privacy, autonomy and fairness (Milano, Taddeo, and Floridi, 2020), and generate risks such as misinformation, filter bubbles and epistemic fragmentation (Milano, Mittelstadt, et al., 2021). RSs have been featured in legal cases involving the endangerment of minors through social media challenges (Edwards, 2022) and the recruitment of terrorists (Centre for Democracy and Technology, 2023): this underlines their deep impact on society. Moreover, according to Pedreschi et al. (2025) RSs are a prominent case of "human-AI coevolution", intended as "a potentially endless feedback loop, wherein users' choices generate data to train AI models, which, in turn, shape subsequent user preferences".

Various attempts to classify the ethical implications (Milano, Taddeo, and Floridi, 2020; Milano, Taddeo, and Floridi, 2021) and the type of influence of RSs (Jesse and Jannach, 2021) have been made, pointing to specific aspects (Helberger, Sax, et al., 2022) or domains (Harambam, Helberger, and Van Hoboken, 2018) of the relationship between algorithmic recommendations and users' agency. This chapter aims to set the philosophical background for the rest of the thesis by outlining how users' agency, intended as their ability to make informed and independent choices regarding access to content on a platform interface, is influenced by RSs. Rather than proposing an exhaustive account of the complex socio-technical intersections underpinning the influence of RSs, I will sketch the theoretical context in which my argumentative line is situated. The innovation brought about by the DSA provisions on RSs lies, at its core, on the principle of user empowerment, which requires platforms to make the user able to exercise an informed agency on the functioning and results of the recommendation process. Despite the skepticism on whether a substantive realization of user empowerment would occur as a result of DSA compliance (Helberger, Van Drunen, et al., 2021) and the discouraging early outcomes in this regard (Fabbri and Boratto, 2025), the fact that this principle is part of the normative foundation of EU dig-

ital regulation allows for the possibility that its practical consequences will eventually materialize. Indeed, the judgment of the District Court of Amsterdam (2025) on how Meta should comply with art. 27 and 38 DSA followed the "underlying purpose" of these provisions, i.e. that "users should have genuine autonomy, freedom of choice and control over the way in which information is presented to them".

The principle of user empowerment emerging from the DSA is situated in a political-philosophical framework pointing to the concept of positive liberty (Berlin, 1969), intended as "the possibility of acting — or the fact of acting — in such a way as to take control of one's life and realize one's fundamental purposes" (Carter, 2022). In fact, online platforms are accessed by individuals voluntarily, so there is no material constraint on the exercise of negative liberty in choosing whether to use their services: if one is not satisfied with the service provided, one can leave it. However, the influence of platforms concerns the type of fruition of the service: if a user wanted to use the service in a different way than the standard one (e.g., a way that respected their authentic self-determination), this user would likely encounter dark patterns or design hindrances, as demonstrated by the District Court of Amsterdam (2025).

The DSA, it can be argued, aims to reinstate users as *naturaliter maiorennes* in Kantian terms (Kant, 2013), i.e. able to make use of their own intelligence to self-determine their online experience. This is not a collectivist form of positive liberty: social movements and collective action against algorithmic oppression are not envisioned (Hampton, 2021; Noble, 2018). It is, rather, an individualist concept of positive liberty, whereby the "government should aim actively to create the conditions necessary for individuals to be self-sufficient or to achieve self-realization" (Carter, 2022). The peculiarity of this conceptual framing seems largely missing from the debates surrounding the implementation of the DSA. To understand how individuals can be effectively empowered through this regulatory approach, it is essential to understand the relationship of influence between RSs and users. To this aim, this chapter first advances an ontological framework to distinguish RSs from other AI systems providing recommendations.Building on this categorization, it characterizes

the relationship between users and RSs as one of reliance of the former on the latter. Subsequently, it proposes an ethical perspective connecting user reliance on RSs to the harms enabled by them, which can be interpreted as systemic risks in the DSA formulation.

2.2 The System - Agent - Patient Framework

The concept of algorithmic recommendation lacks a unified understanding due to the variety of domains in which the corresponding term is used. In fact, if the context of use is not specified, what is referred to as a recommendation includes not only the output of RSs, which may influence users without constraining their freedom, but also the outcomes of decision support systems (DSSs) or automated decision-making systems (ADMSs), whose impact on individuals is direct and often does not depend on their choice. The conceptual boundary between RSs and DSSs has not been clearly established, considering that "there is no standard accepted definition for DSSs" in computer science (Kostopoulos, Davrazos, and Kotsiantis, 2024). Given the wide range of AI systems providing recommendations as the main type of output, drawing clear distinctions among them is complex: the ethical and legal consequences of this definitional gap remain underexplored. Whilst DSSs are "devoted to performing a content-specific task that supports human decision making (although human decisions often tend to be determined rather than supported by it)", RSs "are not content- but context-specific: the content of their output can vary widely depending on the user, but they are directed by a defined aim within a particular context, i.e. maximizing user engagement in a social media platform" (Fabbri, 2023b). Following this argument, if a recommendation always falls under a specific topic within a wider domain (e.g., personalized therapy for lung cancer), then it should represent the output of a DSS. Otherwise, if a recommendation deals with various topics in the same domain (e.g., miscellaneous daily news based on a user's profile), it can be considered the output of a RS.

However, this argument does not provide a boundary that allows to distinguish precisely, from the recipient's perspective, whether a recom-

mendation is the output of DSSs or RSs, as it does not clarify whether the subject that directly faces the implications of the decision can choose whether to follow the suggestion of the system. For example, in the healthcare domain, a recommendation about keeping the appropriate heartbeat will have very different implications if it is produced by a runner's wearable device or by a Holter monitor worn by a patient under anaesthesia in the operating room: although the content of the recommendation is the same, in the former case it is "consumed" by the person directly concerned by it (i.e. the runner whose heartbeat is being measured), while in the latter case it is "consumed" by a third person (i.e. the surgeon) who decides whether it will impact the person directly involved (i.e., the patient).

This example highlights a guiding question for the classification of different concepts of algorithmic recommendation: who has the responsibility and autonomy to decide whether to follow the recommendation? To attempt an answer, I consider three subjects: the system (S), the agent (A) and the patient (P): S is the technology that produces the recommendation; A evaluates the recommendation and decides whether to follow it; P directly bears the consequences of following the recommendation. The relationship between these subjects can determine a taxonomy that allows to distinguish between RSs, DSSs and ADMSs:

- If $A = P$ and $S \neq A$, a RS can be recognised. The subject who receives a recommendation from the AI system and can choose whether to follow it is the same who directly bears the consequences of following it. Therefore, the recommendation can influence but cannot constrain the choices of the subject who is exposed to it (e.g.: a YouTube user sees a list of recommended videos, clicks on one of them and watches it).

- If $S \neq A$, $S \neq P$ and $A \neq P$, a DSS can be recognised. The subject who receives a recommendation from the AI system decides which impact it will have on another person who is not actively involved in the decision-making process but bears its consequences. This is typical of domains where a specific expertise is required, like

medicine or law (e.g.: a judge decides for how many years a culprit should be convicted based on their recidivism score).

- If $S = A$ and $A \neq P$, an ADMS can be recognised. The recommendation coming from the AI system is directly enforced on the subject who must bear its consequences without any human-in-the-loop intervention. In this case, the recommendation de facto becomes an automated decision (e.g.: in the UK in 2020, when final high-school exams could not be taken due to the pandemic, A-level grades, which determine admission to university, were assigned by the so-called Ofqual algorithm without any mediation by teachers or schools (Kolkman, 2020); as the algorithm turned out to be biased, the artificially estimated results were not considered).

The system-agent-patient (SAP) framework could contribute to establishing whether an algorithmic recommendation comes from a RS, DSS or ADMS from the perspective of its human recipient, thereby bringing conceptual clarity on the distribution of responsibility for the output of these AI systems, each of which has different implications on society. In fact, while RSs influence individuals indirectly through nudging mechanisms (Jesse and Jannach, 2021), DSSs and, even more so, ADMSs constrain the autonomy and freedom of the subjects who bear the consequences of following the recommendation but are not responsible for choosing whether to follow it. An objection could be raised against the possibility of overlap between system and agent (S = A), which characterizes ADMSs, as it is in the end a human choice to implement an automated decision without human oversight on its output: therefore, the human agent would exist even if only ex ante. However, it can be argued that such objection does not hold at different levels of abstraction (LoA) (Floridi, 2008) through the example of street-level bureaucratic operations led by AI systems (Gillingham, Morley, and Floridi, 2025): while the decision of adopting an ADMS for a specific administrative task pertains to the LoA of human high-level policy-makers, its application at the LoA of street-level bureaucracy is not mediated by a human agent and, therefore, the decisions on single cases cannot not be pragmatically ascribed

to an identifiable human decision-making process (although high-level policy makers would be at least morally responsible for them). In the case of adopting an ADMS for case handling, the discretionary decision-making power of street-level bureaucrats is in principle transferred to the system's designers (Bullock, 2019), ex ante, and in fact to the system itself, ex post.

2.2.1 Legal applications

The classification based on the SAP framework has the potential to improve the ethical scrutiny and auditing of AI technologies, which are required by European regulations such as the DSA, for what concerns RSs in online platforms, and the AI Act (AIA) (European Parliament and Council, 2024), for all types of recommendations. In fact, while RSs are not explicitly mentioned in the AIA and an interpretation by the European Commission labels them as minimal risk AI applications within this regulatory framework (Commission, 2024a), recommendations "that can influence physical or virtual environments" are included among the output of an AI system according to the definition provided by art. 3(1) AIA [1]. The amendment 740/2023 of the European Parliament proposed to include RSs used by VLOPSEs among the high-risk AI applications of Annex III, but it did not reach the final version of the AIA (Söderlund et al., 2024). Under the DSA, instead, VLOPs have to comply with requirements on the transparency and user control of RSs (art. 27 and 38) and to identify and address the systemic risks enabled by them (art. 34 and 35). Distinguishing between the type of system producing an algorithmic recommendation is therefore necessary to decide which regulatory framework would apply and in which context. Eventually, the two regulations appear inconsistent with regard to risks posed by RSs, which can be minimal according to (a European Commission's interpretation of) the AIA but also systemic according to the DSA.

[1] "'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".

The importance of distinguishing between RSs, DSSs and ADMSs becomes evident considering Annex III of the AIA, which lists the categories of high-risk AI according to domains of application, such as employment, law enforcement and education. The range of applications mentioned in Annex III goes from AI systems used "to place targeted job advertisements" (4(a)), which pertain to RSs, to those aimed at evaluating "the reliability of evidence in the course of the investigation or prosecution of criminal offences" (6(c)), which can be considered DSSs, to those "intended to be used for monitoring and detecting prohibited behaviour of students during tests in the context of or within educational and vocational training institutions at all levels" (3(d)), which are under the scope of ADMSs. In fact, following the SAP framework, it can be argued that: in the first instance, targeted job ads influence, but do not constrain, recipients' ability to apply for specific jobs, as those seeing an ad can freely choose whether to apply for the related job ($P = A$); in the second instance, the subject of a criminal investigation may not have access to the evaluations provided by the system, which are instead used by prosecutors to assess their case ($P \neq A$); in the third instance, the proctoring system adopted to detect whether students cheat yields de facto binding decisions on their conduct during tests which, if not thoroughly reviewed, substitute human examiners' surveillance ($S = A$). While all these AI systems are featured as high-risk AI in the AIA and therefore subject to the same requirements, they have very different consequences on their "patients": in fact, while recipients may choose not to consider a targeted job ad, culprits cannot decide how prosecutors assess the evidence against them with the help of automated evaluations, and students subject to AI-enabled proctoring may not even be able to appeal against a human-made decision on their conduct during tests.

The proposed framework for the classification of automated recommendations between RSs, DSSs and ADMSs – based on the respective coincidence or divergence of system, agent and patient – facilitates the assessment of the ethical and legal implications of these systems in the context of EU digital regulation. As the intersections between the AIA and the DSA have not been clarified yet, fringe cases such as algorith-

mic recommendations may not be adequately assessed by regulators and cause undesirable consequences such as inconsistent compliance patterns, which are already evident in the first algorithmic audits performed under the DSA (Terzis, Veale, and Gaumann, 2024). In fact, some high-risk AI systems according to the AIA, such as those producing targeted job ads, should also be considered within the scope of the transparency and systemic risk provisions for RSs under the DSA (e.g. LinkedIn's RSs). In these cases, the technical documentation required for high-risk AI in the AIA should be complemented by the evidence provided by platforms for the DSA enforcement process. Correctly classifying different concepts of algorithmic recommendation allows to identify cases at the intersection between complementary regulatory frameworks and to clarify the consequent responsibility and compliance burdens for providers. Although not all-encompassing, the SAP framework can facilitate the forthcoming regulatory endeavours by providing a conceptual tool to untangle the complexity of responsibilities related to recommendations.

2.3 An Ethical Perspective on the Influence of Recommender Systems

According to the SAP framework, RSs are recognised when there is an overlap between agent and patient with regard to the recommendation provided by the system: in particular, the agent, who decides whether to follow the recommendation, coincides with the patient, who bears the consequences of following it, while the role of the system is limited to the provision of such recommendation. The responsibility of the decision ultimately lies on the agent, who, in the case of RSs, is often represented by the user of an online platform: for example, on TikTok, the user scrolls through the videos appearing on the interface, decides which ones to watch and ignores the rest, or interacts with the For You Feed in order to find more interesting ones. However, one can argue that, since the RS displays the content in a ranked order, and the user has limited knowledge on the range of available content and on the functioning of the recommendation process, user choices are influenced by the RS (Gillespie,

2014). In other words, deciding to consume a specific content depends also on the presentation of content on the platform interface.

The influence of RSs on users' decisions could be characterized as nudging: according to the original definition in behavioural economics proposed by Thaler and Sunstein (2008), a nudge is "any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives". Since algorithmic recommendations "influence which information is easily accessible to us and thus affect our decision-making processes", they can be interpreted "as digital nudges, because they determine different aspects of the choice architecture for users" (Jesse and Jannach, 2021). Informational nudging through interactive RSs has been proposed as a way to increase user awareness, e.g. by asking users "questions about the contents that are going to be recommended or the categories through which the recommendation is informed" (Fabbri, 2023c). The nudging potential of recommendations may also be used to manipulate users through dark patterns, defined by recital 67 DSA as "practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions".

Following Sax (2021), three conditions need to apply cumulatively for manipulation to occur: first, "the manipulator wants to further her own ends by making use of the other person"; second, the act of manipulating "is by definition intentional on the part of the manipulator"; third, a manipulator attempts to infiltrate someone's decision-making process", often by identifying and exploiting their vulnerabilities, "because the manipulator is reasonably sure that her target is not willing to (fully) cooperate if asked. Empirical evidence is currently insufficient to prove that all these conditions are present in the type of influence exerted by RSs (Jesse and Jannach, 2021). In any case, since technological artifacts do not have intentions, RSs would at most be the means through which manipulation is exercised. This implies that, to argue that users are manipulated by RSs, one should preliminary assume that RSs are intentionally designed in ways that allow the service provider to infiltrate

users' decision-making process in order to steer their choices toward its aims, which do not coincide with the users' aims or to which users do not spontaneously contribute. In these terms, keeping user engagement high is the aim of the provider, which would trick users into spending more time than they otherwise would on its platform to maximise its profit. As intuitive as it may sound, this argument applies better to dark patterns, which entail a specific intentionality of the interface designer that is undesirable for the user (Mathur, Kshirsagar, and Mayer, 2021), than to RSs per se, which are instead aimed at reducing information overload (and may therefore be a desirable platform feature). For this reason, below I propose a theoretical account of the influence of RSs which departs from manipulation.

I argue that the RSs derive their influence from users' dependence on them for accessing information in online platforms and search engines. Therefore, I theorize RSs as *reliance enablers*: recommendations, in both their analogue and digital form, are effective when they can be relied upon, i.e. when their content is considered reliable by a relevant stakeholder. In a user-centred perspective, the influence which the receiver of a recommendation is subject to is the stronger the more reliable is the source from which it comes (as shown also by the case of reference letters for jobs or university applications). As RSs are a means to reduce information overload, the extent to which they can influence an individual's choice depends on how relevant that individual considers the content of the recommendation across time in a specific platform environment. Consequently, the more the users rely upon what is recommended to them without being aware of how they are nudged by RSs, the more they may be vulnerable to the harms that the DSA aims to prevent. By considering RSs as *reliance enablers*, I aim to provide a conceptual foundation for reshaping the role of users as active stakeholders in their interaction with platforms: indeed, if users decide what content to rely on by informing the recommendation process through explicit feedback, their autonomy and online safety could be better preserved.

Reliance is intended here as "as a continued relationship on the basis of one party's dependable habits toward the other" (Deley and Dubois,

2020). I use the concept of reliance instead of trust because I consider the former more relevant to the type of relationship between users and RSs. Indeed, RSs, as a technology, cannot intentionally betray users' expectations on the outcomes of their functioning. A technological artifact does not have the intention to consciously comply with its users' trust mandate (Baier, 1986); it "rather produces dependable habits that we come to rely upon or not" (Deley and Dubois, 2020). The trustees are the designers and owners of such technology, who build and operate it based on aims that might (or might not) conform with the users' mandate. Arguing that the user trusts the RS to suggest relevant content would disintermediate the role of designers and platform owners, which instead shape how users rely on RSs to access information.

It has been revealed that the social network X biases recommendations to prioritize its owner's tweets in each user's For You Tab, regardless of users' interest for them (Clark, 2023): the platform engineers, under the threat of being fired, "deployed code to automatically "greenlight" all of Musk's tweets, meaning his posts will bypass Twitter's filters designed to show people the best content possible" and, as a result, "rank higher than anyone else's in the feed". This example makes it clear that users' reliance on RSs concerns their habitual dependence on algorithmic recommendations that are in turn shaped by the aims of designers and owners, who may or may not fulfil their role of trustees whose priority is user interest.

Trust (and, relatedly, manipulation) emerges in the relationship between those responsible for the design and functioning of a technology and those who use it, rather than directly between the technology and its users. This argument keeps the human in the loop visible by preventing service providers from avoiding accountability for their role of trustees. At the same time, user agency in accessing content is not depicted as dependent on a relationship of trust toward RSs, but rather as enabled by a conscious reliance on RSs. If the recommended content does not satisfy the user, they should be able to intervene on the technology on which they rely in order to opt for a different type of recommendation. In fact, user empowerment is realized not only by choosing whether to consume

a recommended content, but also by deciding how the recommendation process works. In this sense, users can self-determine their online experience by choosing how to rely on RSs.

This argument expands the implications of the SAP framework with regard to freedom: while the coincidence of agent and patient in RSs assigns responsibility for following the recommendation to end users, the influence that the system has on them is not taken into account. Relying on RSs does not hinder users' negative liberty, or freedom from interference, as they can choose at any time to disregard recommendations or leave the platform, as happened with journalists' exodus from X after its takeover by Musk (Barbieri, 2024). However, passive reliance undermines positive liberty: in the majority of platforms, users cannot decide how to steer the system they rely on based on their aims (e.g., through the selection of different options for RSs). Consequently, if the user empowerment principle enshrined by the DSA is not applied, users can exercise negative liberty (i.e. choose to ignore recommended content or leave the platform), but cannot exercise positive liberty (i.e. intervene on how RSs influence their exposure to content).

Empowering users, in this context, means enabling them to shape their reliance on RSs according to their interests and purposes: users should not only be responsible for the choice of following a recommendation, but also for how and which recommendations compose their feed. By mandating transparency and controllability, the DSA aims to transform reliance into a space for positive liberty in Berlin's sense: users are not merely free *from* algorithmic influence, but free *through* it, i.e. aware of how RSs influence them and able to redirect such influence through control options. In fact, passive reliance on RSs fosters the potential of these algorithms to generate harms, which, in the DSA formulation, constitute systemic risks. From this perspective, the dual nature of recommender systems—as both enablers of reliance and sources of systemic risk—arises from the vulnerability of users as the influenced party: once individuals depend on recommendations to access content, they become exposed to the potential harms that such content may inflict on their digital wellbeing and fundamental rights. The empirical

research that could contribute to the understanding of RS-enabled systemic risks in social media platforms is at a relatively early stage, as the DSA mechanism to access non-public data of very large providers has stared materializing since 29th October 2025, while the access to public data already granted by art. 40(12) DSA is actively hindered by platforms (European Commission, 2025d).

2.3.1 Personalization, Epistemic Fragmentation and Harms

Assuming that the harms of RSs stem from users' reliance on them, which represents a source of vulnerability, it is useful to consider how these harms can materialize. Milano, Taddeo, and Floridi (2020) propose an initial taxonomy of the ethical challenges posed by RSs: among the social effects of RSs, they identify a "lack of exposure to contrastive views", giving rise to the so-called filter-bubbles, which can be exploited by manipulative agents in order to increase the frequency with which a content is recommended within specific online communities. As Bozdag and Van Den Hoven (2015) put in evidence, phenomena such as polarization on social media arise because of a subtle manipulation of the contents delivered individually but spread collectively by RSs: through strategic content tagging and by exploiting the networked structure of platforms, political campaigners may be able to redirect public attention on controversial contents which appear on the social media feeds of users. In this regard, Sunstein (2018) has famously pointed out the widespread political implications of digital technologies, including RSs, which have allowed people to "filter what they want to read, see, and hear", not coming "across topics and views that you have not sought out".

In fact, the concept of recommendation is inherently related to that of personalization, although the corresponding phenomena are distinct. In fact, the latter represents the pre-condition for the former. On the one hand, recommendations make sense only if they can be personalized, because, if they were not personalized, they would not be able to reduce the information overload on platforms, which is their main utility for users and providers (Gorgoglione, Panniello, and Tuzhilin, 2019). On the other

hand, personalization can be applied mainly through algorithmic recommendations (in the form of targeted advertisements, suggested contents, etc.): therefore, even if personalization as a design concept makes sense independently of recommendations, its application within the infosphere (Floridi, 2014) often relies on the latter. Therefore, recommendations depend on personalization, whilst personalization is embedded within recommendations from the perspective of its application.

This distinction is required in order to understand how the sociotechnical structure of RSs is related to the epistemic fragmentation of users (Milano, Mittelstadt, et al., 2021), a prominent problem in online platforms. Epistemic fragmentation can be defined as the phenomenon by which individual users lose contact with their peers through online targeted advertising. In particular, as each user is targeted individually by RSs, one cannot know which content another person sees: in this sense, users' knowledge about their common experience on the platform is fragmented, because what they see is the result of personalization and cannot be shared among different individuals. This aspect is even more relevant considering that the effects of personalization do not necessarily imply that each user sees a different array of contents. In fact, an analysis of news recommendations on Google News by Nechushtai and Lewis (2019) found that "users with different political leanings from different states were recommended very similar news".

Epistemic fragmentation is not only a result of the individualization of recommended contents, but it also derives from the opaqueness of the recommendation process, which prevents users from becoming aware of the platform dynamics. This situation can give rise to ethical concerns especially when personalisation is based on implicit user profiling, through which "the system determines what the user is interested in" thanks to implicit data, which include "web usage mining [...], IP address, cookies" and other metadata (Bozdag, 2013). Indeed, if a user is profiled through implicit data, the recommendations will be less transparent and explainable compared to a situation in which "the user customizes the information source himself" (ibidem) by providing explicitly data such as personal interests, demographic information and ratings. In

the context of an epistemically fragmented user experience, the influence of RSs relying on implicit profiling may hold negative aggregate social implications. In fact, when users do not have control over which kind of data is used for their profiling, the recommendations are more likely to bring unwanted contents to their attention.

As a result, users may suffer, on a first dimension, from absolute harms of inclusion or exclusion, which "originate in the nature of the content that is either included or excluded from what is shown to an individual consumer" (Milano, Mittelstadt, et al., 2021): the former occur when genuinely bad and offensive contents (i.e. false claims or racist stereotypes used for promotional purposes) are displayed on the users' profile, whilst the latter occur when essential contents (i.e. important public heath announcements) are omitted, without the user's consent or control on the process. On a second dimension, users can be affected by contextual harms of inclusion or exclusion, which "do not stem from the nature of the content per se, but depend on the context in which the content is delivered" (ibidem): for example, a contextual harm of inclusion may occur when unhealthy food is suggested to obese people or children, who may be more likely to buy them; conversely, a contextual harm of exclusion can be recognised when a job-seeker does not encounter advertisements for positions in their area. The categories of harms produced by RSs do not arise only from implicit signals but may also be a consequence of the data that users choose to provide explicitly. For example, a user may want to provide explicit data about personal unhealthy habits, such as gambling, because they are interested in finding products or offers in the related domain, regardless of their impact on wellbeing. In the same way, some users may give a high rating to recommendations about contents featuring stereotypes that other people may find offensive or unethical: if the latter share interests with the former, they may see such unwanted recommendations due to collaborative filtering algorithms. These cases show that even personalization based on explicit, user-provided signals may originate unexpected harms, which cannot be evaluated just from the point of view of the individual but need to be interpreted within the context of both the platform environment and the

socio-technical structure of RSs.

The influence of RSs also depends on the effectiveness of the recommendation policies implemented in their design, which usually rely on the exploitation or exploration of the space of choices. An exploitative policy aims "to recommend an item that has the highest expected probability of satisfying the user's preferences" (Milano, Taddeo, and Floridi, 2021), whilst an explorative policy focuses on recommending "content with uncertain predicted user engagement for the purpose of gathering more information" about users' interests (McInerney et al., 2018). When RSs rely exclusively on exploitative policies, users can be led into feedback loops that may reinforce their current preferences and lead to bad consumer choices in the long run. For example, a user that usually buys unhealthy food through a delivery app based on exploitative RSs may receive recommendations about the same kind of food every time they want to make an order and therefore their health could be impacted negatively. In this case, an explorative policy could instead propose different kinds of products that do not correspond to the preferences previously expressed by the user, eventually inducing them to find healthier food they like.

Explanations, as a feature of RSs, can impact on users' perception of the recommendation and their subsequent behaviour (Tintarev and Masthoff, 2015). Jesse and Jannach (2021) classify explanations as nudging mechanisms within the Decision Information category based on making information visible. Indeed, viewing explanations can influence user choices and interest in the content they consume. Integrating information about the content within the recommendation could be beneficial for users' understanding of their own preferences. Following the same example as above, a food recommendation might be designed so that the nutritional values of a product that a user has (exploitative policy) or has not (explorative policy) bought before are displayed to them before they can proceed to the order: in this way, the user could be informed about the characteristics of their dietary choices. El Majjodi, A. D. Starke, and Trattner (2022) demonstrate that nutrition labels integrated into RS interfaces act as digital nudges, shifting consumption choices without ma-

nipulation or coercion. Providing explanations would make users aware of the extent to which their preferences have been taken into account by the policy informing the recommendation. For these reasons, explanations for RSs can be considered a kind of pro-ethical informational nudging (Fabbri, 2023c), as they improve user-system interaction in direction of transparency and trustworthiness just through the provision of information. Across these contexts, influence arises not because RSs are inherently manipulative, but because users have no independent reference point outside the system's curated horizon.

2.3.2 From platform to court

The classification of harms presented above covers different cases in which RSs would have a negative impact on users' wellbeing. The wide-ranging implications of harms caused by RSs go beyond the individual, acquiring a societal relevance. A case of absolute harm of inclusion covered by the international press concerns the "blackout challenge" on TikTok, which encourages users to film themselves as they choke themselves to the point of fainting and then regain consciousness on camera: various cases emerged in which minors died while trying the challenge. After recent cases, which happened in the USA (Edwards, 2022) and UK (Sarkar, 2022), some American families decided to sue the platform as it let the challenge spread and target children through its recommendation algorithm (Levenson and Rubin, 2022). While this may at first seem a problem of content moderation, it is, at a deeper level, a consequence of the use of RSs in social media platforms, where their main aim is to increase users' engagement.

As RSs are often based on uninterpretable machine learning models, it might be difficult to attribute the liability for the harm to the platform. In fact, the platform could argue that contents are displayed to users according to recommendation policies that take their preferences into account, so, if the user liked or kept consuming a harmful content which is later reproposed to them, the system should not be blamed. Moreover, as access to the platform by individuals under a certain age should be su-

pervised by parents, it is the parents' duty to control the online activity of their children. To challenge this argument, one should prove that it is the recommendation policy itself to be biased towards contents aimed at maximizing engagement regardless of the vulnerability of the user: according to this perspective, the platform would be liable for designing RSs that influence users' behaviour to fulfil the interests of the system.

A related argument about platforms' responsibility for the content suggested by their RSs is embraced by petitioners in the Gonzalez vs Google case, which deals with "whether Section 230 [of the US Communication Decency Act] shields Google from liability for allegedly recommending ISIS content posted to YouTube to other YouTube users" (Centre for Democracy and Technology, 2023). This lawsuit emerged as a result of the deaths caused by the 2015 terrorist attacks in Paris, France, which were carried out by people recruited by ISIS after being exposed to social media content disseminated by the organization through YouTube RSs. The question posed to the US Supreme Court concerned whether "section 230(c)(l) immunize[s] interactive computer services when they make targeted recommendations of information provided by another information content provider, or only limit the liability of interactive computer services when they engage in traditional editorial functions (such as deciding whether to display or withdraw) with regard to such information" (Supreme Court of the United States, 2022). Petitioners argued that "Section 230(c)(1), which shields intermediaries from liability for publishing third-party content, applies only to claims based on the display of content, not the recommendation of content" (ibidem).

In May 2023, the Supreme Court dismissed the case on the ground that it could not by addressed by antiterrorism law, as the "plaintiffs' complaint seems to fail under [...] our decision in Twitter" vs Taamneh, which concerned the same issue of Gonzalez vs Google (Supreme Court of the United States, 2023). If the Supreme Court's ruling had excluded targeted recommendations from the protection provided by Section 230, implying that "the "recommendation" of content is different from the display of content", platforms would have been forced to change their moderation and recommendation processes and users might have lost

their "rights to like and promote content in forums where they act as community moderators and effectively boost some content over other content" (Ryan-Mosley, 2023). As the long-standing debate between freedom of expression and (online) safety eventually focuses on the impact of the influence of RSs, it is crucial for users to understand how these algorithms function in order to shape their design: the prerequisite for user empowerment is "algorithmic awareness" (Eslami et al., 2015).

2.4 Conclusion

In this chapter, the principle of user empowerment has been considered as the normative paradigm of the EU regulatory provisions on the transparency and controllability of RSs. I argued that this principle is rooted in Isaiah Berlin's account of positive liberty from an individualist perspective: "to create the conditions necessary for individuals to be self-sufficient or to achieve self-realization" (Carter, 2022), the DSA aims to empower users to choose how to inform the recommendation processes shaping their online experience. This interpretation of the DSA requirements for RSs has been adopted by the District Court of Amsterdam (2025), whose judgement has set the ground for future judicial decisions and enforcement actions in this direction.

A preliminary condition to empower users is to account for their relationship of influence with algorithmic recommendations. To do so, I proposed an ontological framework to categorize the different types of recommendations, based on the respective roles of system (producing the recommendation), agent (deciding whether to follow it) and patient (facing the consequences of following it): RSs can be identified when agent and patient coincide, as in the case of platform users who are exposed to content through recommendation and decide whether to consume it. The user is therefore outlined as a passive subject who relies on the RS for their consumption choices. In other words, recommendations enable users' reliance on the system, making them vulnerable to the harms that exposure to certain content can cause. In this context, the application of the DSA provisions on RSs aims foster user agency not only

as consumers but also as citizens that embrace the positive liberty of co-governing the socio-technical infrastructure that influences their choices and experiences. However, for this aim to materialise, user awareness of their possibility to actively control RSs should be increased through algorithmic literacy: indeed, users who understand how feeds are curated interpret content exposure more critically and feel more in control of their information-consumption habits (Eslami et al., 2015).

Chapter 3

Regulatory background

This chapter presents the EU legal sources on which the regulation of recommender systems is based. It reviews relevant provisions of the Digital Services Act, cognate regulation such as the Code of Conduct on Disinformation, the AI Act and the GDPR. The focus is on whether and how the enforcement of EU digital regulation can mitigate the unfair consequences of the influence of online platforms on users through RSs. As a result of this analysis, I argue that, to enforce the principle of user empowerment, platforms should be required to: 1) include granular explanations for RSs on the interface, not just in terms and conditions, to foster users' awareness; 2) provide users with appropriate control options to intervene directly on the mechanisms through which RSs influence them.

3.1 Introduction

As discussed in Chapter 2, RSs influence the main aspects of the user experience and can give rise to ethical concerns related to privacy, autonomy and fairness (Milano, Taddeo, and Floridi, 2020), to name but a few. The political economy of platforms based on profiling and recommendations has been notably addressed by Zuboff (2019) with the concept of "surveillance capitalism". However, independent research and auditing on the design and functioning of the RSs of online platforms has been

historically prevented by proprietary constraints.

There is a normative discrepancy between the widespread use of RSs in various domains and the methods through which their ethical and societal impact can be evaluated. Transparency demands have become increasingly pressing, as the implementation of opaque models may have problematic consequences on the users' ability to retrieve relevant information and define their online identity. As algorithmic recommendations often rely on implicit personal data, such as browsing and click-through history, and their functioning is not explained to users, their influence is not accountable. Although explanations for RSs have been addressed by research in Explainable AI (Tintarev and Masthoff, 2007), their effects on algorithmic design and on the different stakeholders within the recommendation process have not been extensively assessed. Moreover, even when explanations and control options are provided on real-world platforms, users are not able to interact explicitly with them, apart from providing limited feedback. The limitations regarding the transparency and controllability of automated recommendations are supposed to be addressed by the provisions of the EU regulations like the DSA, which require online platforms, including marketplaces and social media, to explain the functioning of RSs and allow users to shape their design. However, the effectiveness of the application of these regulatory provisions will depend on the extent to which people understand how RSs work and how they can influence their functioning.

In this chapter, I focus on whether and how the new European regulatory context around RSs can address the risks and opportunities stemming from this pervasive digital technology, especially from the perspective of mitigating the unfair consequences of the imbalance between platforms' influence and users' self-determination. To do so, I analyse the implications of the DSA provisions on RSs transparency and user control within the broader context of EU digital regulation, including the AI Act and the General Data Protection Regulation. As a result of this analysis, I argue that, in order for the aims of the DSA provisions about RSs to be fulfilled, the principle of users' self-determination needs to be substantiated by: 1) easy accessibility of explanations on how the rec-

ommendation process works and how users can be influenced by it; 2) an extended possibility for users to intervene directly on the strategies through which RSs target them on the platform's interface.

3.2 Digital Services Act (DSA): from recommender systems to systemic risks

The DSA is the first supranational regulation addressing the transparency and controllability of RSs with the aim of empowering users of online platforms (Fabbri, 2023a; C. Starke et al., 2025a). It defines RS as "a fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service or prioritize that information, including as a result of a search initiated by the recipient of the service or otherwise determining the relative order or prominence of information displayed" (DSA, art. 3 (s)). This definition highlights the method ("fully or partially automated"), aim ("to suggest"), content ("specific information"), target ("recipients of the service"), input ("as a result of a search initiated by the recipient") and output ("determining the relative order or prominence of information displayed") of a recommendation process.

From the perspective of the transparency and user control of RSs, art. 27(1) requires platform providers to explain "in their terms and conditions, in plain and intelligible language, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters". The parameters considered must include, at least, "the criteria which are most significant in determining the information suggested to the recipient of the service" (content) and the reasons for its "relative importance" (ranking) (art. 27 (2)). Additionally, when options to modify or influence the main parameters are mentioned in the terms and conditions, platforms should provide, in correspondence with the list of ranked recommendations, a "directly and easily accessible" functionality "that allows the recipient of the service to select and to modify at any time their preferred option" (art. 27 (3)).

Moreover, "providers of very large online platforms and of very large online search engines that use recommender systems shall provide at least one option for each of their recommender systems which is not based on profiling"[1] (art. 38). It is worth noticing that, while the provisions of Article 27 apply to all online platforms, the application of Article 38 is limited to VLOPSEs, which therefore represent the only environments in which users will always have the option to choose between at least two types of recommendations[2].

These provisions aim to foster users' empowerment through their awareness of how the recommendation process works and their direct intervention on the platform's interface. Their rationale is to "ensure that recipients of their service are appropriately informed about how recommender systems impact the way information is displayed, and can influence how information is presented to them" (recital 70). A right to explanation for RSs could be identified in this formulation: in fact, the "easily comprehensible manner" of presenting the parameters of RSs so that "the recipients understand how information is prioritized for them" (recital 70) can come to effect only if RSs are explainable[3].

The paragraphs of Article 27 aim to address four of the aspects of the definition of RS provided by Article 3(s): method, target, input and output. In particular, as a result of the enforcement of the DSA, the traditionally passive role of the target might be reversed, as the recipient could determine the method (through the choice of parameters) and, indirectly, also the input (the type of data to be processed through the parameters) that the RS would use to produce its output. This opportunity to enhance transparency and users' self-determination has not been welcomed by a prominent digital company like Meta, which has stated that "the breadth of some of the auditing obligations under the DSA should be clarified/improved as these could become a barrier for growth in the

[1] Profiling is defined here according to Article 4 (4) of the GDPR.

[2] It is plausible to state that all the VLOPSEs identified by the European Commission (`https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops`) use profiling for automated recommendations, so the provision of Article 38 applies to all of them.

[3] The right to information outlined here is mirrored by Article 13-15 of the GDPR, which will be considered later.

sector" (Meta, n.d.).

However, online platforms that are not VLOPs or VLOSEs using RSs based on profiling will not be obliged to provide options for users to modify or influence the parameters if this possibility is not specified in the terms and conditions, and platforms arguably have no interest in providing this possibility voluntarily. Therefore, Article 27 formally grants users the right to influence the recommendation process but only in some limited cases which may not be likely to happen, as (Helberger, Van Drunen, et al., 2021) point out. Moreover, the practical impact of these provisions will depend on how the terms with which the legal text refers to the technical features of RSs ("parameters", "option", "criteria", "functionality") will be interpreted by platforms and regulators. Further clarifications on how to apply these provisions can be extrapolated from the legal intersections analysed below.

3.2.1 Regulatory Intersections on Transparency and User Empowerment for Recommender Systems

A promising development in this regard is the approval of the Code of Conduct on Disinformation (CCD) (European Commission, 2025b) by the Commission and the Board of Digital Services on 13 February 2025. The CCD is the result of the integration of the 2022 Strengthened Code of Practice on Disinformation within the DSA compliance framework. As a result of this process, the CCD "will become a relevant benchmark for determining DSA compliance regarding disinformation risks for the providers of VLOPs and VLOSEs that adhere to and comply with its commitments" (European Commission, 2025h). Five VLOPs (Facebook, Instagram, TikTok, YouTube, LinkedIn) and two VLOSEs (Bing, Google Search) are among the signatories of the Code: hence, for these providers, applying the CCD commitments will be taken as evidence of compliance with the related DSA provisions. This is highlighted by the Commission's opinion on the CCD, according to which "Certain Commitments (such as for example Commitment 19 on transparency of recommender systems) are drafted in a language that mirrors that of certain provi-

sions of Regulation (EU) 2022/2065. When this is the case, the Code of Practice on Disinformation represents a further operational and complementary layer, including reporting obligations, on top of existing legal obligations" (European Commission, 2025c). Commitment 19 requires "Signatories using recommender systems [...] to make them transparent to the recipients regarding the main criteria and parameters used for prioritising or deprioritising information, and provide options to users about recommender systems, and make available information on those options". This Commitment explicitly requires signatories to provide options for users to control RSs and, as the first benchmark for compliance with the DSA provisions on transparency and controllability of RSs, offers a grounded perspective on how to interpret the requirements of art. 27: VLOPs that signed the CCD, apart from disclosing the main parameters of their RSs, must provide control options beyond the possibility of disabling profiling-based recommendations as per art. 38.

While the DSA does not define what the main parameters of RSs are, this concept is not new in EU law, as it appears for the first time in Regulation 2019/1150 (European Parliament and Council, 2019), referred to as Platform to Business (P2B), in the context of product ranking in search engines and marketplaces. In particular, art. 5 P2B requires providers of "online intermediation services" to "set out in their terms and conditions the main parameters determining ranking and the reasons for the relative importance of those main parameters" (1) through "publicly available description, drafted in plain and intelligible language" (2). As can be seen, there is a close overlap between art. 5 P2B and art. 27 DSA. The most noticeable differences are the use of the term "ranking" instead of "recommender systems" and the fact that transparency is assumed to the benefit of business users rather than consumers. In fact, according to Recital 24 P2B, "The specific design of this transparency obligation is important for business users as it [...] should help business users to improve the presentation of their goods and services, or some inherent characteristics of those goods or services". Ranking can be considered an output of the recommendation process, as the definition of RS in art. 3(s) DSA highlights: without making any reference to art. 5 P2B, the DSA

borrows the concept of ranking parameter to frame the RSs transparency obligations for online platforms, which are clearly directed at end users, as Recital 70 DSA details.

Moreover, to "facilitate the compliance of providers" (art. 5(7) P2B), the Commission published guidelines on ranking transparency pursuant to art. 5 (European Commission, 2020), which offer the regulator's interpretation on how to identify main parameters following a so-called "right level of detail" approach. On the one side, "the description to be provided has to go beyond a simple enumeration of the main parameters, and provide at least a 'second layer' of explanatory information", such as "the company-internal 'thought process' that was used for identifying the 'main parameters', as a way to also derive the 'reasons for their relative importance' (guideline 22). On the other side, "an excess of information can mean that, in effect, no meaningful information is provided to users", who should therefore not be overwhelmed "with too lengthy or complicated descriptions", also to "avoid the risk of enabling the deception of consumers or consumer harm" (guideline 25). In practice, to identify the main parameters, "providers could consider what drove the design of the algorithm in the first place" (e.g. "a desire to ensure that consumers found goods or services that were local, cheap, of high quality"), "the top type of result on its service", or "what would make the services' consumers most satisfied [...], for example by considering why it chose the filters that are offered to consumers to sort results on its services". These guidelines remain purposefully vague on the level of detail of the information that should be provided to business users, also because "providers of online search engines shall [...] not be required to disclose algorithms or any information that, with reasonable certainty, would result in the enabling of deception of consumers or consumer harm through the manipulation of search results" (art. 5(6) P2B). According to this perspective, too detailed information may be used to reverse-engineer ranking results and manipulate consumers' choices: therefore, if this provision has to be understood as a point of reference for the scope of art. 27 DSA for compliance, the Commission may not expect platforms to publicly disclose their algorithmic metrics and weights under

the DSA. Nonetheless, since, with its proposed Digital Omnibus (European Commission, 2025g), the Commission aims to repeal the P2B regulation, its Article 5 and the related guidelines on ranking transparency may become dead letter.

Instead, which type of options shall be included by platforms on their interface has been partly clarified by the Commission's guidelines on the protection of minors under art. 28 DSA (European Commission, 2025f). In particular, options should allow minors "to select content categories and activities they are most or least interested in including explanations [...] during the account creation process and regularly throughout the minor's time on the platform" and should directly influence the recommendations provided by the system (guideline 67(c)). Moreover, platforms shall "Meaningfully explain why each specific piece of content was recommended [...], including information about the parameters used and the user signals collected for that specific recommendation" (guideline 67(b)): this clarifies the type of explanations that are expected in art. 27(1) and (2) DSA. Apart from the provisions directly complementing art. 27 DSA, platforms shall "Prioritise 'explicit user-provided signals' to determine the content displayed and recommended", where 'explicit user-provided signals' refer to "user feedback and interactions that indicate users' explicit preferences, both positive and negative, including the stated and deliberative selection of topics of interest, surveys, reporting, and other quality-based signals"[4] (guideline 65(g)). Platforms are also encouraged to "provide information and nudge minors toward searching for new content after a certain amount of interaction with the recommender system" (guideline 66(b)) to avoid *rabbit holes* (i.e. "the phenomenon where users are continually served content that leads them deeper into a specific topic or viewpoint, often becoming more extreme or polarised in the process" (Panoptykon Foundation, Irish Council for Civil Liberties, and People vs Big Tech, 2023)): this represents the first-ever call by the EU regulator to platforms to nudge a category of users in

[4]Reporting includes feedback "such as "Show me less/more", "I don't want to see/I am not interested in", "I don't want to see content from this account," "This makes me feel uncomfortable," "Hide this," "I don't like this," or "This is not for me."", which VLOPs like Facebook list as options to control their RSs within the remit of art. 27 DSA (Meta, 2024b)

a certain direction.

The Code of Conduct on Disinformation also requires signatories to "give users the option of having signals relating to the trustworthiness of media sources into the recommender systems" (Measure 22.2) and publish "information outlining the main parameters their recommender systems employ in this regard" (Measure 22.3). These requirements aim to detail the user control provisions for RSs in the direction of the trustworthiness of recommended content. The responsibility of defining and incorporating trustworthiness signals into platforms' RSs might set new scenarios for the online information ecosystem: in practice, major platforms would have to publicly state their trustworthiness indicators for content and how these are used as signals to feed RSs. No VLOP has apparently implemented these provisions as of now.

The European Media Freedom Act (EMFA) (European Parliament and Council, n.d.) advances relevant requirements for the source of recommended content. While not directly related to RSs, in fact, its art. 3 introduces the "right of recipients of media services to have access to a plurality of editorially independent media content", also in online platforms. According to its art. 18, media service providers can self-declare to VLOPs, based on criteria of editorial responsibility and independence, to enjoy additional safeguards against content moderation practices such as visibility restrictions and demotion through RSs. Moreover, the European Board for Media Services "shall regularly organise a structured dialogue between providers of very large online platforms, representatives of media service providers and representatives of civil society" in order to "foster access to diverse offerings of independent media on very large online platforms" (art. 19(1) EMFA).

These requirements outline principles of diversity and authoritativeness for the provision of media content in VLOPs, complementing the user empowerment and trustworthiness perspectives emerging from the DSA and CCD. These principles, considered together, can substantiate meaningful personalization in the access to information, by increasingly giving users control over what they can, and therefore may want to, see online. The achievement of meaningful personalization in the context

of the DSA depends, however, not only on the existence of legal provisions, but also on their interpretation and application. While the DSA enforcement process lacks transparency and therefore accountability due to the almost complete unavailability of relevant legal documents (Fabbri, 2025b), much of the leverage for a substantive implementation of the regulatory requirements on RSs is currently left to the benchmarking process developing between VLOPs, independent auditors, and the Commission.

3.2.2 Advertising as a Form of Recommendation

In the context of advertising, the DSA requires online platforms to disclose to users "meaningful information directly and easily accessible from the advertisement about the main parameters used to determine the recipient to whom the advertisement is presented and, where applicable, about how to change those parameters" (art. 26 (1d)). Moreover, VLOPs that display advertisements shall "compile and make publicly available in a specific section of their online interface, through a searchable and reliable tool that allows multicriteria queries and through application programming interfaces, a repository" (DSA, art. 39 (1)) featuring the following information:

> "(a) the content of the advertisement, including the name of the product, service or brand and the subject matter of the advertisement; (b) the natural or legal person on whose behalf the advertisement is presented; (c) the natural or legal person who paid for the advertisement, if that person is different from the person referred to in point (b); (d) the period during which the advertisement was presented; (e) whether the advertisement was intended to be presented specifically to one or more particular groups of recipients of the service and if so, the main parameters used for that purpose including where applicable the main parameters used to exclude one or more of such particular groups; (f) the commercial communications published on the very large online platforms [...]; (g) the total number of recipients of the service reached and, where applicable, aggregate numbers broken down by Member State for the group or groups of recipients that the advertisement specifically targeted." (art. 39 (2)).

The first four points of the cited paragraph concern the metadata of

the advertisement: its content, who paid for it, the duration of its permanence on the platform. According to point (e), the platform is required to indicate whether the advertisement was targeted and, if so, the main parameters used for including or excluding categories of users from the targeted. Point (g) would allow to understand indirectly the correspondence between specific clusters of users and the advertisement by which they have been targeted in each EU country. The enforcement of this article has the potential to address the epistemic fragmentation of users due to online targeted advertising considered by (Milano, Mittelstadt, et al., 2021). Indeed, if users can access a public repository with information about the parameters used by platforms to segment them into groups for targeting purposes, they can have an idea of how many other people see a particular advertisement and why they see it. The access to this information can reduce the individualization and fragmentation of online experience, as users could eventually become aware of collective platform dynamics, although probably not at a very granular level.

3.2.3 Dark Patterns on Platform Interfaces

As the main driver of what is showed on platform interfaces, RSs can also generate "dark patterns", intended as "practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions" (recital 63 DSA). These practices include "presenting choices in a non-neutral manner, such as giving more prominence to certain choices through visual, auditory, or other components", "repeatedly requesting a recipient of the service to make a choice where such a choice has already been made [...] or making certain choices more difficult or time-consuming than others" (ibidem). According to the recent ruling by the District Court of Amsterdam, the second instance of this list of dark patterns can be identified in the option to disable profiling-based RSs on Instagram and Facebook, as outlined in Chapter 1.

According to art. 25 (1) DSA, "Providers of online platforms shall not design, organise or operate their online interfaces in a way that deceives

or manipulates the recipients of their service or in a way that otherwise materially distorts or impairs the ability of the recipients of their service to make free and informed decisions". The Commission has not yet issued guidelines on how art. 25 applies to specific practices, including "(a) giving more prominence to certain choices when asking the recipient of the service for a decision; (b) repeatedly requesting that the recipient of the service make a choice where that choice has already been made, especially by presenting pop-ups that interfere with the user experience; (c) making the procedure for terminating a service more difficult than subscribing to it" (art. 25 (3)). Therefore, the role of RSs in connection with dark patterns still has to be, at least legally, clarified.

3.2.4 The Systemic Risk Framework

The provisions outlined above are part of a wider regulatory scope. Section 5 of the DSA, including articles from 33 to 43, is dedicated to the obligations for providers of very large online platforms and search engines - VLOPSEs with a shortened acronym - to manage what is referred to as systemic risks. In the DSA, systemic risks stem from the design, functioning and use of VLOPSEs and do not concern all the other intermediaries. Online platforms and search engines should be designated as VLOPSEs if their active monthly user base corresponds to at least 10% of the EU population, or 45 million people (art. 33). Systemic risks should be identified by VLOPSEs (art. 34) and mitigated through specific measures (art. 35). However, the DSA does not provide an explicit definition of systemic risks, which are only specified in a list including:

> "(a) the dissemination of illegal content through their services;
>
> (b) any actual or foreseeable negative effects for the exercise of fundamental rights, in particular the fundamental rights to human dignity enshrined in Article 1 of the Charter, to respect for private and family life enshrined in Article 7 of the Charter, to the protection of personal data enshrined in Article 8 of the Charter, to freedom of expression and information, including the freedom and pluralism of the media, enshrined in Article 11 of the Charter, to non-discrimination enshrined in Article 21 of the Charter, to respect for the rights of the child enshrined in Article 24 of the Charter and to

> a high-level of consumer protection enshrined in Article 38 of the Charter;
>
> (c) any actual or foreseeable negative effects on civic discourse and electoral processes, and public security;
>
> (d) any actual or foreseeable negative effects in relation to gender-based violence, the protection of public health and minors and serious negative consequences to the person's physical and mental well-being." (art. 34 (1)).

After systemic risks have been identified, the measures that VLOPSEs should adopt to mitigate them include, according to art. 35 DSA, adapting the design and functioning of their services, their terms and conditions, content moderation processes, algorithmic systems, advertising systems, reinforcing their internal risk management processes, cooperating with trusted flaggers and other providers, and flagging deepfakes, among other things.

The lack of an explicit definition of what constitutes a systemic risk in the DSA determines the need for a case-by-case reasoning considering the socio-technical affordances of platform design and architecture, including: "(a) the design of their recommender systems [...]; (b) their content moderation systems; (c) the applicable terms and conditions and their enforcement; (d) systems for selecting and presenting advertisements; (e) data related practices of the provider" (art. 34(2)). The crucial role of this provision is highlighted by the fact that, among the first DSA provisions to be challenged by a VLOP, i.e. Amazon, at the Court of Justice of the EU there is Art. 38 (Beatriz Botero Arcila, 2024), which requires VLOPSEs to "provide at least one option for each of their recommender systems which is not based on profiling" and Art. 39, which obliges them to compile a publicly accessible repository on their advertising practices. Platforms' design features which enable systemic risks also support their business model, so VLOPSEs have the incentive to try and comply minimally with provisions that challenge their profit while requiring additional resources for compliance.

The DSA outlines two approaches to generate evidence to support the enforcement of its systemic risk framework: one internal to the Commission, based on the collaboration between the European Centre for Al-

gorithmic Transparency (ECAT), at the Joint Research Centre (JRC), and the DSA enforcement team at the Directorate-General for Communications Networks, Content and Technology (DG Connect); one external to the Commission, based on digital social research produced through the data access mechanisms based on Art. 40. In particular, access to non-public platform data is granted only to vetted researchers through an application to the relevant national Digital Services Coordinator (DSC). According to recital 13 of the Commission's delegated act on data access for vetted researchers, based on art. 40(4) DSA (European Commission, 2025e):

> "Data that can be requested in order to study systemic risks or their mitigation in the Union may evolve in the future. Current examples of such data include data related to users of the services, such as profile information, relationship networks, individual-level content exposure and engagement histories; interaction data such as comments or other engagements; data related to content recommendations, including data used to personalise recommendations; data related to the targeting of advertisements and profiling, including cost per click data and other measures of advertising prices; data related to the testing of new features prior to their deployment, including the results of A/B tests; data related to content moderation and governance, such as data on algorithmic or other content moderation systems and processes, including changelogs, archives or repositories documenting moderated content, including accounts as well as data related to prices, quantities and characteristics of goods or services provided or intermediated by the data provider."

3.3 The Absence of Recommender Systems in the AI Act: a Missed Opportunity?

The provisions of the DSA fill the gaps of the EU Artificial Intelligence Act (AIA) concerning RSs: in fact, as outlined above, RSs are not explicitly mentioned in the AI Act and an interpretation by the European Commission labels them as minimal risk AI systems (Commission, 2024a). References to automated recommendations appear in a few cases without being explained. The first occurrence is the definition of AI system as

"a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, *recommendations*, or decisions that can influence physical or virtual environments" (art. 3(1) AIA). Subsequently, recommendations are mentioned in the context of high-risk AI systems (art. 7(2d), human oversight (art. 14(4b)), testing of high-risk AI systems in real world conditions outside regulatory sandboxes (art. 60(4k); art. 61(1d)). In all these instances, "automated recommendations are considered from the perspective of the outcome and not of the process: therefore, they are merely regarded as outputs of an AI system that can have an impact on human decision-making, whilst a specific focus on the design principles of RSs and the risks posed by their biases is completely lacking" (Fabbri, 2023c). This choice may appear inconsistent with the widespread impact that algorithmic recommendations have on users, which can also include serious harms.

The amendment 740/2023 [5] Parliament, 2023 of the European Parliament proposed to include RSs used by social media VLOPs among the high-risk AI applications of Annex III, but it did not reach the final version of the AIA. That amendment would have made the AI component of RSs part of the high-risk AI applications listed in Annex III (1(8(ab))) as "AI systems intended to be used by social media platforms that have been designated as very large online platforms [...] in their recommender systems to recommend to the recipient of the service user-generated content available on the platform". The final version of Annex III mentions only one type of RSs, namely AI systems used "to place targeted job advertisements" (4(a)). This means that RSs are not excluded from the scope of the AIA, but their inclusion as a high-risk AI systems concerns only specific applications, such as those relating to jobs. However, the fact that the expression "recommender systems" does not appear in the AIA text makes it difficult to interpret the extent to which this regulation intertwines with the DSA in regulating RSs.

[5]The complete list of amendments can be found at: `https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html`

The reasons for which RSs in general have not been considered a high-risk AI technology in the final version of the AIA concern the fact that recommendations impact indirectly rather than directly on individuals. A comparative example might be helpful: automated credit risk assessment, which has been included in Annex III since the beginning of the draftingn phase, is supposed to output a score that helps human decision-makers determine whether a client is suitable to receive a loan. In this case, the system is devoted to performing a content-specific task that supports human decision making (although human decisions often tend to be determined rather than supported by it). Algorithmic recommendations, instead, are not content- but context-specific: the content of their output can vary widely depending on the user, but they are directed by a defined aim within a particular context, i.e. maximizing user engagement in a social media platform. For this reason, the recommendation does not raise ethical concerns per se, but as regards its domain of application.

In the domain of RSs, the intersection of the AIA and the DSA creates complex regulatory challenges, particularly visible in cases like deepfakes generated by general-purpose AI systems spreading through VLOPs' recommendations. While the DSA identifies VLOPs as environments enabling systemic risks through specific socio-technical features, the AIA identifies general purpose AI models with high capabilities (art. 51 AIA) as technologies producing systemic risks per se. This distinction between enabling and producing systemic risks reflects the different regulatory focuses: the societal implications of content diffusion in the DSA versus the direct effects of large-scale technical capabilities in the AIA. However, both frameworks ultimately face the common challenge of requiring providers to assess broadly defined societal risks subject to significant value-laden disagreements—a challenge that cannot be resolved through technical solutions alone.

3.4 General Data Protection Regulation (GDPR) and the Right to Explanation

Article 22 of the General Data Protection Regulation (GDPR) (European Commission, 2016) addresses "automated individual decision making, including profiling" stating that "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her" (art. 22(1)). RSs are based on profiling, so they can be considered within the regulatory scope of this article. However, there are three exceptions to the provision reported above, which "shall not apply if the decision: (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent" (art. 22(2)). When exceptions (a) and (c) apply, "the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision" (art. 22(3)). Moreover, according to the fourth paragraph of the article, sensitive data should never be collected for profiling. However, it often happens that sensitive data are inferred from non-sensitive data which act as proxies: for instance, income level could be inferred from household address.

Exception (a) could be claimed in all the cases in which users are asked to accept the terms of service of a platform, which define the contract between the data subject and the data controller. Exception (c) applies when the user is asked for online consent, for example for what concerns cookies. Therefore, it can be argued that automated recommendations comply with the GDPR requirements, given that, when accepting the terms of service, the user is often giving consent to profiling and inferences. On the one side, Article 27 of the DSA aligns with the rationale

of GDPR by requiring that explanations of RSs are presented in the terms and conditions, which are not often read by users and therefore may not impact on their awareness of their rights. On the other side, Article 38 of the DSA complements the GDPR by requiring that very large online platforms keep a repository of targeted advertisements, so that users can view the outcome of legitimate profiling.

The nudging potential of automated decision-making systems may, in some cases, lead humans to conform uncritically to their assessments, thereby making the application of Article 22 of the GDPR controversial (Lagioia and Simoncini, 2018). In fact, the safeguards against decisions that do not involve humans in the loop are not clarified in Article 22, which does not state how users can determine whether a decision is completely automated. Instead, a hint in this direction is provided by articles 13 and 14, on the right to information, and 15, on the right to access, according to which the controller must give information about the existence of automated decision-making, including profiling, as referred to in Article 22, and, at least in such cases, meaningful information about the logic used, as well as the significance and the intended consequences of such processing for the data subject (Lagioia and Simoncini, 2018). This is complemented by Recital 71, which suggests that profiling "should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision". The right to explanation envisaged here is crucial to substantiate the safeguarding claims of the citepd articles, but it is not described in further detail.

This lack of precision has been criticized by Wachter, Mittelstadt, and Floridi (2017), who identify "several reasons to doubt both the legal existence and the feasibility of such a right": in fact, "the GDPR only mandates that data subjects receive meaningful, but properly limited, information (Articles 13–15) about the logic involved, as well as the significance and the envisaged consequences of automated decision-making systems". Moreover, "the ambiguity and limited scope of the 'right not to be subject to automated decision-making' contained in Article 22 (from

which the alleged 'right to explanation' stems) raises questions over the protection actually afforded to data subjects" (ibidem). The DSA goes in the direction of implementing the right to explanation outlined in the GDPR, but the effectiveness of explanations in enhancing users' autonomy is still debated. Future empirical research should be aimed at establishing whether the presence of explanations would substantially contribute to substantiate the users' rights envisioned by these regulations.

3.5 Conclusion

Algorithmic recommendations determine not only what we see on platforms, but also our potential interest for new or different categories of content. This influencing potential can be interpreted as an instance of the "new emerging grey power" of tech companies, which "is exercised about which questions can be asked, when and where, how and by whom and hence what answers can be received in principle" (Floridi, 2015a). A platform like TikTok, which is mainly managed through RSs, is a prominent example of this tendency: as the interface is based on an endless flow of recommended content through which the user scrolls, the contents that the user ends up seeing more frequently are related to the single videos that they watch for a longer time. This exploitative policy has already caused harm because, if a video on which a vulnerable person casually spends a few seconds concerns a dangerous activity, then that individual will see the same content more and more and may eventually be influenced by it. In this sense, platforms control the questions that users pose about their interests and, subsequently, the answers that they get: in this way, digital companies end up informing a substantial part of users' online, and sometimes offline, experience.

Explanations may be a countermeasure to this harmful impact of recommendations, as they have the potential to make users aware of some of the questions that platforms shape for them. I argue that, in order for this potential to be realized, explanations and control options should be integrated as a readily available, standard feature of recommendations which people may choose to use when they want to, or that appear as a

pop-up on the interface of online platforms as a form of nudging. Thanks to such a policy, users could understand why they receive specific recommendations, become aware of the extent to which they are influenced by RSs and, subsequently, decide to redirect this influence towards aims set by them instead of platforms. This is a necessary condition for users' ability to control RSs explicitly through direct interventions on the interface: indeed, explanations should complement granular and proportional control options, as argued in the next chapter.

The DSA requires digital companies that use RSs and targeted advertising to build mechanisms to grant transparency and controllability to users, in order to enhance their self-determination and understanding of the systems they use. However, if users are not interested in receiving explanations or using control options, the provisions of the DSA may not have the expected results. As (Tintarev and Masthoff, 2012) underline, "the explanations affect a user's mental model of the recommender system, and in turn the way they interact with the explanations". Even a focus on explorative rather than exploitative recommendations may not be sufficient to limit the negative outcomes of pervasive RSs. Indeed, from the perspective of digital companies, exploration is mainly a means to get to know users even better than they currently do, by gathering data on unexplored fields of potential interests and preferences (McInerney et al., 2018). This can lead to an even deeper nudging, which is realized through incremental exposure to contents that can provide fine-grained information on how to induce users to like what they do not know they like yet. This is why the guidelines for protection of minors under art. 28 DSA, for the first time, hint at a government-mandated nudging performed by platforms on their interface to prevent harmful outcomes such as *rabbit holes*. How this nudging would be performed is left to platforms' implementations, but the path indicated by the regulator clearly aligns user empowerment through granular explanations and direct control with an explicit intention to steer users' attention away from patterns perceived as dangerous for them. The platforms that signed the Code of Conduct on Disinformation will also have to implement trustworthiness signals into their RSs, which first requires them to declare what they con-

sider as trustworthy sources.

I argue that an effective right to explanation and user control is a preliminary condition for users' self-determination in the platform environment. Explanations can be considered a means to mitigate the negative consequences of the power imbalance between platforms and users. Users cannot shape automated recommendations according to their interests and needs without firstly knowing how and why they are targeted and influenced by RSs: in fact, if someone doesn't know how a system works, they are unlikely to be able to make that system work better. At the same time, users should be enabled to directly steer the recommendation process in the direction of their best interest, In this regard, my contribution points to a prominent policy problem: as explanations are the building blocks of transparency, in order to support self-determination through transparent recommendations it is firstly necessary to educate users to understand not only "what recommenders recommend" (Jannach et al., 2015), but also why they recommend what they recommend. If it is not properly met by regulators on time, this sociotechnical requirement may constrain the positive ethical and societal impact of the DSA provisions. In conclusion, I think that, in order to reduce the power imbalance between platforms and users and limit the influence that the former exert on the latter, policy-makers should: 1) enforce explanations as a user-friendly tool to foster awareness that users can experience on the interface and not only read in the terms and conditions; 2) grant users the possibility of intervening directly and substantially on the strategies through which RSs target them on the platform's interface.

Chapter 4

Auditing Recommender Systems for User Empowerment in Very Large Online Platforms under the DSA

This chapter evaluates the state-of-the-art of VLOPs' compliance with the DSA provisions on RSs through an analysis of the audit and systemic risk reports of Instagram, TikTok and YouTube. What emerges from the reports is that platforms fail to provide substantive interpretations of the legal requirements and align towards minimal compliance. Building on these results, I outline a perspective for the future of VLOPs' RSs grounded in speculative design. Meaningful personalization should integrate algorithmic choice, balancing proportionality and granularity in RS customization, and content curation, ensuring diversity and authoritativeness to mitigate systemic risks. By bridging legal analysis, platform governance and user-centered design, this contribution outlines how platforms can move beyond minimal compliance toward a model that prioritizes user empowerment and content pluralism.

4.1 Introduction

As highlighted by the previous chapter, policymakers' attention has been drawn to the potential risks of RSs, such as their supposed detrimental impact on minors (European Commission, 2024d), their influence on voting intentions (Commission, 2024b), and the challenge of assessing the truthfulness of conflicting news (Vincent and Gismondi, 2021). However, there is no conclusive evidence that RSs, as an element of platform design, directly cause these effects (Sala, Porcaro, and Gómez, 2024; Marciano et al., 2025). Whether the measures introduced by EU law can effectively address the challenges posed by RSs remains open to debate. However, this uncertainty points to the problem of how legal obligations should be designed to enable meaningful user control while ensuring that the implementation of these obligations by platforms meets societal expectations. This chapter explores the interplay between the DSA, its application by very large online platforms (VLOPs), and the design of RSs by addressing the following research questions:

RQ1. *What do the EU legal requirements for RS transparency and control entail, and how have they been implemented by selected VLOPs?*

RQ2. *Mediating between regulatory principles and platforms' practices, how can users be empowered, during their interaction with a RS, to achieve a meaningful personalization of their online experience?*

In this study, I aim to provide a perspective on how legal, technical, and ethical dimensions may converge to shape the future of personalization in online platforms. I examine the compliance strategies of three social media VLOPs—TikTok, Instagram, and YouTube—by reviewing their DSA audit and systemic risk assessment reports. This analysis highlights key gaps between regulatory expectations and platform practices, presented in Section 4.2. Then, I propose a speculative design framework for RS governance in Section 4.3, outlining design features that could foster meaningful personalization while enhancing users' autonomy and control over recommendations. Finally, I integrate insights from legal, technical, and ethical considerations to outline actionable pathways for

aligning regulatory goals with technical affordances and discuss future directions in Section 4.4.

4.2 Analysis of the implementation

As outlined in the previous chapter, the obligations on the transparency and user control of RSs are mainly found in art. 27 DSA. However, the lack of definitions of the technical terms in this article, coupled with their varied (and too often convenient) application by VLOPs in the first round of audit and systemic risk reporting, leaves the interpretative ground open to uncertainty. Some audit reports indeed provide hints on how the terms of art. 27 have been interpreted by VLOPs. According to Instagram's DSA audit (EY, 2024a), redacted by EY, main parameters should be intended as "the top Predictive Events and their corresponding top Signals, with a Predictive Event being the output of a machine learning model that predicts a probability that some event or action will occur, and Signals being the raw, feature-level inputs that feed into the predictive model and determine what the predictive event will be for each user". The same document describes the most significant criteria as "top predictive events and their corresponding top signals", thereby repeating the definition of main parameters. Given the regulatory uncertainty in this regard, and the necessity of an analytical standard, I propose a tentative breakdown of the terminology of art. 27 in Table 1.

Building on the previous legal analysis of the EU regulation supporting the transparency and user control of RSs, I now examine its implementation by three social media VLOPs mainly focused on short-video content: Instagram, TikTok, and YouTube. These platforms have been chosen for two main reasons. Firstly, these are among the most used platforms for content discovery by young users and, through their short-video format, reinforce a flow experience that may stimulate addictive behavior (Montag et al., 2019; Qin, Omar, and Musetti, 2022), even if this has been contested (Marciano et al., 2025). Secondly, and relatedly, these have been a target of the European Commission investigations under the DSA concerning RSs, especially regarding their impact on mental well-

Term	Proposed Interpretation
`Parameters`	The signals used by the system to select and rank candidate recommendations (e.g., watch time).
`Criteria`	The metrics underlying how certain information or parameters are prioritized in the recommendation process (e.g., the relevance of watch time in selecting and ranking candidate recommendations).
`Options`	The available control features provided to users to influence or modify the parameters (e.g., the possibility to select different levels of content sensitivity).
`Functionalities`	The specific interface elements that allow users to choose an option to influence the system's parameters (e.g., selectors for levels of content sensitivity).

Table 1: Key terms of art. 27 and their proposed interpretation.

being and political influence.

Instagram is undergoing infringement proceedings for the design of its interface, "which may exploit the weaknesses and inexperience of minors and cause addictive behavior, and/or reinforce so-called 'rabbit hole' effect" (Commission, 2024c). TikTok is subject to three proceedings, which concern, on the one side, its "persuasive design patterns", considered "problematic for the physical and mental well-being" of users (Commission, 2024d), and, on the other side, the role of its RSs in facilitating "coordinated inauthentic manipulation or automated exploitation of the service", supposedly linked to the recent Romanian presidential elections case (Commission, 2024b). Finally, YouTube, while not currently undergoing proceedings, has received a request of information from the Commission in October 2024 regarding the "parameters used by" its RSs, "as well as their role in amplifying certain systemic risks, including those related to the electoral process and civic discourse, users' mental well-being (e.g., addictive behavior and content 'rabbit holes'),

and the protection of minors" (European Commission, 2024f). The compliance with the provisions on RSs is therefore one of the main areas in which these VLOPs are being held accountable under the DSA.

I carried out an analysis of the implementation of the DSA requirements on RSs based on three DSA-related documental sources for each of the selected VLOPs: the audit report, the audit implementation report, and the systemic risk assessment report. The independent auditors, EY for Instagram and YouTube and KPMG for TikTok, mostly relied on a walkthrough approach (Light, Burgess, and Duguay, 2018) to highlight potential mismatches between the documentation provided by VLOPs and the actual platform design. After reviewing the parts of the aforementioned documents referring specifically to art. 27 and 38 DSA, I integrated the analysis of auditors' findings and platforms' statements with the additional documentation available in the VLOPs' Transparency Centers, to which some reports refer. Based on the evidence collected through this review, I performed an ad-hoc walkthrough on the mobile interface of each VLOP to verify whether the publicly available information provided in compliance with the DSA corresponds to the options and functionalities accessible through the platform's interface. Due to space limitations, my analysis cannot granularly consider all relevant aspects of the aforementioned reports, but it will highlight the information that is fundamental to assess VLOPs' compliance with the DSA provisions on RSs (art. 27 and 38) and their implications for systemic risks.

4.2.1 Instagram

Audit report. The outcomes of Instagram's audit (EY, 2024a) are positive for all the paragraphs of art. 27 and 38. The audit mostly relies on Meta's system cards (Meta, 2024b), which describe Facebook and Instagram RSs parameters and control options. In the case of Instagram, ten RSs are described (Feed, Stories, Explore, Reels chaining, Search, Feed recommendations, Suggested accounts, Notifications, Comments, Threads Feed), along with the signals influencing each of them and the related control options for users. The reported signals are 442, while the control options

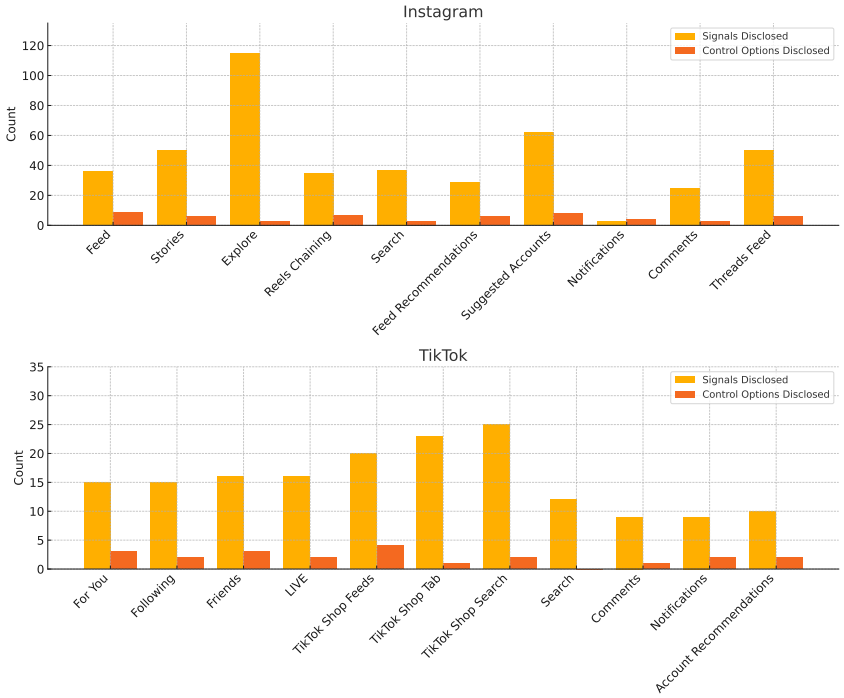


Figure 1: Comparison of signals and control options disclosed in Instagram's and TikTok's audit reports, broken down by RS type.

55. Figure 1 shows the breakdown of signals and controls for different RSs types. Where "relative importance" as per art. 27(2) is intended to mean the relative weight of single parameters in influencing the recommendations, the cards do not explain the relative importance of parameters, nor the reasons for their relative importance: therefore, under this interpretation, Instagram would not be compliant with art. 27(2b), although the audit reasonably assures otherwise, maybe based on a different unstated interpretation. Moreover, not all the control options available are described: indeed, in the control dashboard for reels, there is an option enabling users to choose which amount of political content from people they do not follow they want to see (more, standard, less), but this option is not stated in the relevant system card (Feed recommendations Manage suggested content). Given the positive assessment of all the aspects related to art. 27 and 38, Instagram's audit implementation report (Meta, 2024a) does not address these.

Aside from compliance considerations, it can be observed that the signals disclosed are both implicit (e.g., watch time) and explicit (e.g., like), with a prevalence of the former. The control options, apart from those allowing the disabling of personalization as per art. 38, which result in a ranking of posts based on reverse chronological order, are mostly related to traditional user interaction features, such as *save, mute/unmute, follow/unfollow, hide, report*. I argue that these options do not allow users to intervene directly on which category of content is prioritized, thereby preventing them from effectively steering recommendations towards their preferences. The options aimed at providing more direct controllability are available just for content recommended from accounts that a user does not follow and include: *interested/not interested, specific words and phrases*, to hide recommended content containing such expressions in hashtags or captions, *change sensitive content* and the above mentioned *change political content*, both allowing users to choose which amount of content classified as "sensitive" or "political" they will be exposed to (more, standard, less are the options). The rationale for setting control options with different granularity according to the specific RS is not explained: in practice, only Feed recommendations can be controlled

through more sophisticated approaches (e.g., *change sensitive content*), even though the impact of selecting the related options on the underlying algorithm is not explained.

Systemic risk assessment report. As regards Instagram's systemic risk report (Meta, 2024c), Meta firstly highlights that its RSs "are designed to try to prevent the recommendation, recirculation, or amplification of potentially policy-violating or otherwise problematic content". The VLOP actively practices demotion as a means of policy enforcement: "using my repository of signals, my algorithms may down-rank content for a variety of reasons and do not recommend certain accounts, or content, such as those related to repeat policy violators in feeds". Meta has identified four domains of intervention that would help identify and address the systemic risks enabled by its RSs: "Cross-functional Integration", i.e., "the ability to ingest signals from cross-functional teams at scale to help inform enforcement against violating content in a consistent manner"; "Abuse of Hashtag Feature", i.e., "preventing manipulation of hashtags intended to distribute policy violating content"; "Recommendation Features", as recommendations of accounts to follow "could connect threat actors to minors"; and "Content Ranking", whose risk of leading users to exposure to harmful content is reduced "by limiting the role of shares and comments in the distribution of sensitive topics". Apart from fostering user empowerment through the explanations in the RSs cards, Meta claims to offer "users the ability to control how much of certain types of content (including sensitive or low quality content) they see in their feeds", even if no such thing as a "low quality content control" can be found either in the system cards or on the platform. The report does not provide any details on how the technical features of Instagram's RSs are tied to specific systemic risks and which related interventions Meta has designed.

4.2.2 TikTok

Audit report. According to its Help Center (TikTok, n.d.), the TikTok RSs include For You, Search, Notifications, Account recommendations,

Following feed, Friends tab, LIVE feed, and TikTok Shop, which recommends commercial products through some of the preceding RSs. Also the STEM feed (TikTok, 2024c), which prioritises educational content related to scientific subjects, should be added to these, even if it is not mentioned in the documents reviewed. The main parameters to predict which content a user might like are: user interactions (e.g., likes, comments, shares, watch time, and other engagement metrics), content information (e.g., hashtags, captions, and soundtracks), and user information (e.g., profile details and account settings) (TikTok, 2024c). Each of these main parameters is explained through the signals relevant to it: for example, for the For You Feed (FYF), the signals coming from user interactions include "content you like, share, comment on, and watch in full or skip, as well as accounts of followers that you follow back" (TikTok, n.d.). The options to modify or influence the main parameters primarily enable users to exclude or include specific content from their feed, encompassing, again according to the Help Center: *not interested, refresh your feed, filter keywords, unfollow, manage notifications*, *filters* and *sort by* (only for Shop Search), *unsync your contacts* and *remove suggested accounts* (only for Account recommendations). The signals disclosed are in total 170, while the control options 22. The breakdown of signals and controls for different RSs types is shown in Figure 1. The assessment of compliance with art. 27(2) is positive in TikTok's audit report, although the same consideration made above for Instagram regarding the weight of parameters applies here. However, the report, drafted by KPMG, negatively evaluates the implementation of art. 27 (1) and (3) and 38 and draws a more substantial interpretation of their requirements (KPMG, 2024). In fact, regarding the negative outcome for both art. 27(1) and 27(3), KPMG "identified that the design of the controls in place and their respective control objectives were not sufficient and appropriate to meet the Specified Requirements". The main issue raised concerns the "options to modify or influence those main parameters" (art. 27(1)), which are not deemed sufficient for the following reasons:

> "1. options must enable users to influence or modify the 'main parameters', as defined by TikTok (user information, content informa-

tion, interaction information) and therefore affect the information that will be presented to the user; and 2. options must be a tool or a feature which the user engages with, over and above a user interaction (defined as a main parameter and already informing the user in the Help Center article that how they use the app / interact with content, will impact what they see) [...] 3. the following controls are not considered to be 'options to modify or influence those main parameters':

1. Limiting pool of content: these are content controls that TikTok offers to users that restrict certain content from being retrieved by the recommender systems but do not 'modify or influence' the main parameters (listed above) of the recommender systems.
2. Non-personalized options: these offer users separate experiences that are not based on profiling.
3. Settings: these are user preferences that affect the user's experience on the Platform generally (like language and translation preferences and location services)".

These motivations attempt to provide a consistent definition of the scope of the terms used by art. 27 and 38, introducing the concept of option as "a tool or a feature which the user engages with, over and above a user interaction", which is instead one of the main parameters for TikTok. Options should give users a substantial possibility to influence the main parameters and should be made accessible through the functionality of art. 27(3). For example, according to this perspective, an option should not consist in allowing users to simply disable watch history in the settings of their profile, but instead in granting them the right to enable or disable the use of their watch history as a parameter for specific types of RSs at certain intervals and with respect to certain content categories. KPMG considers the non-personalized options provided according to art. 38 as outside the scope of the "options to modify or influence those main parameters": while the non-personalized options are indeed options, their exclusion from the scope of art. 27 according to KPMG might be due to the fact that they are already covered by art. 38 and so not sufficient for substantially complying with the former provision. The negative outcome of the audit for art. 38 amounts to the observation

that, for the Stories RS, the non-personalized option was not available during part of the evaluation period: a less prominent problem than the previous one.

Systemic risk assessment report. For what concerns RSs, the systemic risk report of TikTok (TikTok, 2024b) focuses in particular on the protection of minors and information integrity in the context of the For You Feed (FYF), adopting measures such as "banning certain hashtags and/or blocking searches associated with emerging risks": as a result, harmful content may be either demoted or unavailable in FYF. TikTok offers users "tools to diversify the content displayed, to understand why videos have been recommended, to choose that certain keywords will not be displayed to them and to reset their FYF as if they were new to TikTok": these explanations and control features are at the center of the platform's efforts to tackle systemic risks emerging through RSs by empowering users to control the dissemination of both theirs and others' content. The connection between these design principles and the technical affordances of the platform is not clarified, and this is reflected by the negative audit result related to the control options described above. In the audit implementation report (TikTok, 2024a), TikTok does not specify how it will practically address the shortcomings of the control options provided for its RSs from a socio-technical perspective; instead the VLOP declares that "it will complete its assessment of relevant processes and systems to identify any risks to compliance objectives that address the existing options, and, as appropriate, implement new controls or strengthen existing controls over these risks". This sentence is repeated throughout the whole report, and not just for the provisions regarding RSs.

4.2.3 YouTube

Audit report. The audit report of YouTube, drafted by EY, aligns with that of Instagram (EY, 2024a): the auditor confirms compliance with all the provisions of art. 27 and 38 (EY, 2024b) and, for this reason, YouTube's audit implementation report does not provide further information in this regard (Google Ireland Limited, 2024). Unlike Instagram

and TikTok, YouTube's audit report does not mention different RSs but focuses on listing a series of high-level parameters and options. The former mostly focus on user interaction data and include: "watch history, search history, channel subscriptions, watch time, sharing, likes/dislikes, video clicks, user selection of "Not interested" and "Don't recommend", survey responses, and Google Activity". The latter are mostly aimed at granting users' control of the data available for feeding recommendations and include:

> "Remove, pause, or delete specific videos from watch history, and/or searches from search history through the Google account, Turn off or delete watch and search history, Ads personalization and additional settings that may impact data available for recommendations such as Choose topics in a user's recommendations on the Home Page and Watch page, Remove recommended content from homepage, Clear "Not interested" and "Don't recommend channel" feedback, Manage recommendations and search results on Google Account activity, Manage liked videos (i.e., remove liked videos on YouTube), Subscribe and Unsubscribe to YouTube channels and Delete survey answers".

While YouTube's parameters are more than 80 billion according to the VLOP's 2024 transparency report under the Strengthened Code of Practice on Disinformation (Google, 2024a), the areas in which they have been grouped are 12 and there is a one-to-one matching between the parameters and the related control options. The options allowing to remove videos from, or outright delete and turn off, watch history, search history, and Google account activity are taken as evidence of compliance with art. 38. Therefore, differently from Instagram and TikTok, YouTube does not have a single specific functionality for disabling profiling-based recommendations, but it allows the user to do that through individual interventions on various parameters. The only control option allowing users to actively choose the content of recommendations is *Choose topics in a user's recommendations on the Home Page and Watch page*, which consists in the selection of topic tags "based on your existing, personalized suggestions" and "on content related to what you interact with", as explained by YouTube Help (YouTube Help, 2024b). This option is complemented

by YouTube's search filters, which, apart from providing standard filtering tools (based on chronological order, view count, length of a video) also include tags allowing users to rank search results based on their previous interaction with the video (such as "watched" and "unwatched") (YouTube Help, 2024a). This is the only example of personalized search filter found in the VLOPs considered. The other control options rely on a negative feedback from the user (e.g., *Not interested, Don't recommend channel*).

Systemic risk assessment report. According to YouTube's systemic risk report (Google, 2024b), while "Poorly designed or controlled recommender systems may increase the risk that harmful content goes viral", "properly functioning recommender systems should decrease risk by increasing the visibility of high-quality and trustworthy content and by promoting a diversity of topics and sources for users to explore". This high-level statement hints at YouTube's view on DSA-oriented recommendations, which can mitigate, or at least not enable, systemic risks. In particular, the proper functioning of RSs (which, in this context, may equate to their alignment with regulatory principles) depends on their design and control and results in the promotion of authoritative and diverse content. As can be seen, these principles reflect those emerging from my previous legal analysis: on the one side, user empowerment through control over the design of RSs (as per the DSA and the CCD) and, on the other side, authoritativeness and diversity (as suggested by EMFA). However, the report does not mention how YouTube would operationalize these principles into the functioning of RSs. It does, however, underline that one of the roles of RSs is to support content moderation by "elevating high-quality and trustworthy content", which is considered "a more proportionate approach to addressing harmful content risk than removing content altogether", as this "can present risks to freedom of expression and information".

4.2.4 Comparing the Emerging Evidence

Differing interpretations. The emerging evidence regarding RSs governance in the three VLOPs considered above allows us to draw preliminary considerations on both auditing and compliance practices. First of all, the fact that the auditor of TikTok is different from that of Instagram and YouTube invites us to compare how they respectively inspected the platforms' algorithms. Both EY and KPMG adopted a walkthrough approach to verify whether the parameters, options, and functionalities of RSs described in the platforms' documentation were consistent with what users can find while using their interfaces. It is not stated whether such a walkthrough would be performed on the app or desktop version of the platform, nor which phases it entailed at a granular level. As (Light, Burgess, and Duguay, 2018) present it, the walkthrough method consists of "systematically and forensically step[ping] through the various stages of app registration and entry, everyday use and discontinuation of use", and allows to collect "a foundational corpus of data" that can be used to assess an "app's intended purpose, embedded cultural meanings and implied ideal users and uses". In the case at hand, the walkthrough merely focused on the technical design features of the VLOPs, leaving societal and user-centric considerations in the background.

On the legal side, however, EY and KPMG adopted different perspectives. Indeed, while EY positively assessed compliance regarding all the DSA provisions on RSs for Instagram and YouTube, KPMG did the opposite for TikTok, except for compliance with art. 27(2). The main point here concerns the interpretation of which control options should be made available for RSs according to art. 27(1) and (3). EY reasonably assured that the options and functionalities available were sufficient for compliance with art. 27, while KPMG took the opposite stance, considering that options to limit the pool of recommended content, non-personalised options and settings are outside the scope of art. 27. In fact, since the disclosed options for controlling RSs differ only marginally between Instagram, YouTube, and TikTok, their positive or negative evaluation, it

can be argued, mainly depends on the interpretation of what substantive control options would entail. In particular, the three platforms have standard options that allow to modify user profile and engagement information, including *unfollow, dislike, not interested* associated with content to be excluded from the recommendation process. What is lacking, with respect to the array of parameters disclosed, are the options allowing users to actively determine their exposure to different types of recommended content: in this perspective, only *change sensitive/political content* on Instagram, the STEM feed and *refresh your feed* on TikTok, followed by the selection of topics in the initial preference elicitation, and Youtube's tags allowing users to explicitly include specific topics in the recommendations go in the direction of empowering users' active choices. However, for what concerns Instagram, the impact of selecting "more", "standard" or "less" for the *change sensitive/political content* options on the inner functioning of every RS is not explained. As regards TikTok, the possibility of only starting from scratch by refreshing the feed clashes with users' interest in keeping some aspects of their feed that would be lost. Lastly, for YouTube, users cannot autonomously input their topics of interest but just select tags within a limited and opaquely generated list.

The emerging market of algorithmic audits. The motivations behind different auditing practices, and the related compliance efforts by platforms under the DSA, have been addressed by (Terzis, Veale, and Gaumann, 2024), who argue that "the breadth of systemic risks that the DSA requires to be considered also allows considerable room for framing, bringing the auditor's priorities and worldview and constructing the DSA requirements around them, rather than the other way around". The auditors' priorities also include maintaining good business relationships with VLOPs that may already use their services for financial audits, as the Delegated Regulation on the Performance of Audits (DRPA) (European Commission, 2023b) clarifies that concerns for conflict of interest "should not exclude auditing organisations who have performed statutory financial audits" (recital 8 DRPA). While "responsibility for the initial formulation of the benchmarks against which compliance is/will be sought, rests with the audited organisation (DRPA art. 5 (1) (a))", the

"back-and-forth between the auditing organisation and the [...] audited organisation" may "impact on the way auditing benchmarks are formulated and/or standardised in the future" (Terzis, Veale, and Gaumann, 2024). Consequently, given that, de facto, "the primary responsibility for the formulation of benchmarks is left to the auditing organisation, [...] benchmark disparities amongst different auditors may incentivise platforms to choose their assessors and auditors based on their benchmarks (easy or difficult, simple or complicated)" (*ibidem*).

Based on these considerations, it can be argued that the market of algorithmic auditing for RSs is going to expand but not necessarily by diversifying its offer: it might indeed align along minimalist interpretations of art. 27, causing the usual concern of transparency washing (Zalnieriute, 2021). Benchmarking disparities regarding which control options are needed to comply with art. 27 may not be solved unless a conscious regulatory intervention, i.e., through guidelines or standards, addresses the terminological and design ambiguities found in the DSA provisions on RSs. Moreover, the partial public disclosure of audit and systemic risk reports has not substantially clarified how platforms (plan to) mitigate the harms emerging from the influence of RSs and whether concrete socio-technical interventions are envisioned to substantiate the user empowerment enshrined in the EU regulatory framework. The next yearly round of DSA-related audits will shed light on whether the risk of minimal compliance persists and more substantial steps towards clear benchmarking have been taken by VLOPs in the meantime.

4.3 Speculative Design for User Empowerment

The prevalent paradigm in RSs is to optimize based on user feedback, without requiring any direct user intervention. However, several studies suggest that users benefit (e.g., through better recommendation acceptance) when they have some degree of control over the system (O'Donovan et al., 2008; Bostandjiev, O'Donovan, and Höllerer, 2012a; Harper et al., 2015; Bart P. Knijnenburg, Bostandjiev, et al., 2012; Sun et al., 2023). So, there seems to be some space for designing controllable RSs, where the

user may tune the RS's output via some interface. The design of such an interface involves several key decisions, including the selection of appropriate UI elements (Dokoupil, Boratto, and Peska, 2024a; Millecamp et al., 2018), the granularity of user responses (Cena et al., 2017), and the most effective way to label user choices (Amoo and Friedman, 2001). A substantive implementation of art. 27 and 38 DSA (Fabbri, 2023b) would require VLOPs to put in place effective options for users to intervene on the algorithmic parameters to change the output of the recommendations. The complexity of these options should be gradual enough to meet the heterogeneity of skills across users (Jin, Cardoso, and Verbert, 2017). At the low level, users would be able to give feedback on the recommendations they receive, on its topic and on the content creator (like/dislike) and could see explanations in discursive or graphic form (e.g., word clouds). At the higher level, users could be allowed to modify the degree of personalization of the recommendations, e.g., by choosing the percentage of personalized recommendations they want to see. They should also be able to decide which types of their data, including profile information and interaction signals, can be used as input for RSs.

Existing work on RSs controllability has primarily focused on user interface affordances and transparency mechanisms (Parra and Brusilovsky, 2015; Dokoupil, Boratto, and Peska, 2024a). However, these efforts often remain constrained by the current platform-centric model, where transparency and user control are treated as auxiliary features rather than fundamental design principles. In contrast to this, I adopt a speculative design approach (Auger, 2013; Baumer, Blythe, and Tanenbaum, 2020) that integrates proportionality, granularity, diversity, and authoritativeness as core components of RS architecture. The role of speculative design is "first, to enable us to think about the future; second, to critique current practice" (Auger, 2013): in the context of my work, this approach allows to overcome the aforementioned issues in DSA implementation with actionable principles for meaningful personalization through transparent and controllable RSs. By shifting the focus towards user-driven re-weighting and exposure-aware ranking, my framework envisions a recommendation process where personalization is not just a

system-imposed feature but a user-driven mechanism for content curation. Therefore, I propose four design perspectives to enable meaningful personalization and I group them into two broader categories: (1) *enabling algorithmic choice*, which includes proportionality and granularity, and (2) *directing content curation*, which includes diversity and authoritativeness. I also highlight directions for interdisciplinary research bridging regulatory incentives with design innovation in RSs.

4.3.1 Enabling Algorithmic Choice

This category concerns mechanisms through which users can influence *how* the recommendation process works.

Proportionality refers to the ability of users to control the extent to which personalization influences their recommendations. Users should be able to adjust the degree of algorithmic intervention in ranking models, ensuring that personalized content aligns with their preferences without fully overriding serendipitous discovery. In practice, this could be implemented via simple UI controls like sliders: for instance, a user might reduce the proportion of exploitative recommendations to discover new content. Prior research has explored interactive mechanisms to let users dial such trade-offs (Sun et al., 2023; Harper et al., 2015). For example, allowing users to manipulate a *personalization–diversity slider* was found to significantly affect their satisfaction and perceived recommendation quality (Sun et al., 2024). In a field experiment with an interactive movie RS, users with different exploration preferences were given interfaces to control how broad or narrow their recommendations were; interestingly, the level of exploration control a user chose was strongly predictive of their eventual satisfaction with the system (Sun et al., 2024). Moreover, a longitudinal study on music recommendations showed that when users are given a personalization slider for their playlist, their chosen personalization level can evolve over time: an initially high personalization setting might be gradually reduced as the user seeks more variety (Liang and Willemsen, 2022a). This suggests that users value being able to dynamically re-balance personalization as their needs change.

Future RSs could allow users to maintain multiple recommendation profiles, dynamically re-weighting their preferred level of personalization across different contexts (e.g., a highly personalized profile for news and a less exploitative profile for exploratory browsing).

Granularity concerns the level of detail at which users can modify RS parameters or signals. Fine-grained control means users can adjust individual system features at different levels of depth, and immediately see how these adjustments impact their recommendations. Instead of treating the RS as a black box, platforms could provide controllable dials for key factors, e.g., a user could increase the weight of *topic relevance* while decreasing that of *recently viewed* items. Users should be able to inspect which features (e.g., engagement metrics such as number of likes, content attributes like genre or source) are driving the recommendations and then modify them. The idea of actively shaping the outcome of RSs in this way resonates with users' intentions (Kluttz and Mulligan, 2019; C. Starke et al., 2025a). Prior studies have accordingly investigated various interfaces for feature-level control. For example, TasteWeights (Bostandjiev, O'Donovan, and Höllerer, 2012b) allowed users to visually adjust the weights of different influencing factors in a hybrid music recommender. In the movie domain, a plug-in system called MovieBrain added interactive settings like genre-based filters and personalized hybrid weighing, which improved users' perception of the recommendation quality (Dooms, Pessemier, and Martens, 2014). Relatedly, researchers have developed UIs for interactive feature weighing where users can tune the importance of content attributes and immediately see the adjusted recommendation list (Parra, Brusilovsky, and Trattner, 2014; Millecamp et al., 2018; Sun et al., 2023; Liang and Willemsen, 2022b). Recent work on multi-objective RSs also shows that letting users adjust the importance of objectives (e.g., relevance vs. novelty) can improve their satisfaction and understanding of the system (Dokoupil, Peska, and Boratto, 2023; Dokoupil, Boratto, and Peska, 2024a). In essence, granularity turns RS control into a two-way conversation: the system offers suggestions while the user can fine-tune the dials (feature weights, filters, etc.) to clarify their intent. Such transparency and con-

testability at the feature level do not only improve the explainability of results but also ensure that users with diverse needs feel empowered by the RS rather than constrained by it (C. Starke et al., 2025a). While major platforms have started providing more granular controls thanks to regulatory incentives, their efforts remain quite limited (Help, 2025).

4.3.2 Directing Content Curation

This category focuses on the *scope* and *quality* of user intervention, determining *what* aspects of content the user can influence.

Diversity refers to mechanisms that enable users to explore various perspectives in their feed, reducing echo chamber effects while maintaining relevance. To come out of echo chambers, users should be able to adjust how diverse or novel their recommendation lists are. One approach is to provide a *diversity slider*, coupled with the personalization control mentioned above, so that users can explicitly balance highly focused recommendations with more eclectic ones. Studies have found that users who had greater access to diversity settings showed higher satisfaction when those controls aligned with their personal exploration preference (Sun et al., 2024; Liang and Willemsen, 2022a). Moreover, allowing users to create multiple preference profiles (for different genres or contexts) has been suggested as a way to increase diversity (Reviglio and Fabbri, 2024): e.g., a user could maintain one profile for news content and another for entertainment, effectively preventing any single profile from over-personalizing and thereby preserving a wider content spectrum. However, users' perception of diversity can be tricky, as highlighted by recent research that revealed that users tend to favor intuitive diversity metrics (such as diversity in genres or source types) over abstract algorithmic diversity measures like dissimilarity-optimized ranking (Dokoupil, Boratto, and Peska, 2024b). Additionally, users quickly reach a point of saturation: beyond a certain level, small increases in objective diversity are not even noticed (Dokoupil, Boratto, and Peska, 2024b). This aligns with observations that users often perceive diversity in non-linear ways; a moderate jump from very narrow to somewhat

varied content yields a big increase in perceived diversity, but further broadening beyond the user's interest zone may decrease returns or even confuse the user (Dokoupil, Boratto, and Peska, 2024b). Furthermore, an important paradox emerged in a study of multi-objective recommendations: although many users claim they want novel and diverse items, their actual interactions often gravitate toward the familiar (Dokoupil, Peska, and Boratto, 2023). This suggests that simply injecting diversity is not enough; the system should also help users make sense of novel content. Effective design could include showing explanatory cues or grouping diverse items with brief descriptions, so that users understand why those less familiar recommendations appear (Dokoupil, Peska, and Boratto, 2023). By introducing user-driven diversity controls, alongside supportive explanations, RSs can enable exploration while maintaining a sense of relevance, allowing users to deliberately broaden their horizons without feeling lost or mis-targeted.

Authoritativeness concerns emphasizing *trustworthy*, reliable sources in recommendations, which is a crucial aspect in domains like news, health, or any context where disinformation is a risk. This perspective gives users (and, by extension, the system) the ability to prioritize content based on credibility signals. RSs can integrate these preferences by incorporating source credibility metrics into the ranking process. For example, a platform might allow users to activate a "credibility boost" functionality which, when enabled, ranks items from reputable outlets higher in their feed. Internally, this could be implemented by increasing the weight of a source quality feature in the ranking algorithm. Recent research prototypes have demonstrated how multi-objective ranking can incorporate ethical or quality-oriented constraints without severely degrading user experience by treating the ranking task as a constrained optimization problem, where the maximization of predicted user engagement is subject to constraints on diversity, fairness, or credibility of content (Tkachenko, Dhaouadi, and Jedidi, 2022). While such optimization can be computationally heavy, new methods have been proposed to predict the optimal trade-offs much faster, enabling these credibility and quality considerations to be applied in real time (Tkachenko, Dhaouadi,

and Jedidi, 2022). From a user's perspective, authoritativeness controls might also be presented in more transparent ways: for example, showing trust labels or source ratings alongside content, and giving users the option to filter or sort content by these indicators. User studies indicate that when users are aware of the trustworthiness of content (through labels or explanations), they often appreciate the ability to favor trustworthy sources, especially in high-stakes contexts like health information retrieval (Tkachenko, Dhaouadi, and Jedidi, 2022). Enabling users to incorporate authoritativeness signals into RSs and clearly explaining them aligns with the requirements of Measures 22.2 and 22.3 of the CCD, as outlined in Section 2. Ultimately, by allowing users to easily elevate authoritative content (or demote dubious content), future RSs can serve both individuals' priorities and societal interest, fostering an ecosystem where quality information is more readily discovered.

4.3.3 Proposing a Compliance Standard

The legal analysis in Section 2 outlined how user empowerment through explanations and effective control, in DSA and CCD, and diversity and authoritativeness in content curation and moderation, in CCD and EMFA, are fundamental principles of the emerging EU regulatory framework around RSs. Such principles have been confirmed also by the analysis of VLOPs' reports in Section 3, which, despite highlighting diverging interpretations, underscored the need for substantive user control options (KPMG, 2024) and the promotion of trustworthy and diverse content as a means to combat disinformation (Google, 2024b). These principles are aimed at realizing meaningful personalization, intended as the conjunction of users' awareness and active involvement in informing recommendations.

The four design perspectives introduced in this section can represent a framework for platforms' compliance with the EU legal requirements on RSs: moving from opaque, one-size-fits-all algorithms to transparent, user-steerable ecosystems where people can co-curate the content they consume. As mentioned above, one of the main problems regard-

ing the evaluation of how transparent and controllable recommendations should look like stems from the absence of guidelines on how to interpret and operationalize the related regulatory requirements. The design framework I propose would allow users and regulators to evaluate the transparency and controllability of RSs according to proportionality and granularity, from the perspective of algorithmic choice, and diversity and authoritativeness, from the perspective of content curation. Platforms' RSs could be ranked through a standardized score measuring the extent to which users can influence their functioning across these four perspectives, resulting in an easily accessible visualization that compares the different RSs used by each VLOP. To connect my design framework with an initial standard for platforms to implement DSA-oriented RSs, I propose four checkpoints corresponding to each aspect outlined in my design framework.

1. **Proportionality**: *To what extent can users decide the proportional impact of different RS types, policies and data on the algorithmic logic?*
2. **Granularity**: *How deep and specific can users' intervention on algorithmic parameters and weights be?*
3. **Diversity**: *Which metrics are adopted to determine how diverse the recommended content is and to what extent can users intervene on them?*
4. **Authoritativeness**: *Based on the platform's definition of credible and trustworthy sources, can users decide how frequently they want to interact with authoritative content?*

4.4 Conclusion

Understanding and controlling the recommendation process in social media is a complex endeavor, in which those who succeeded the most are now identified as VLOPs under the DSA. With its requirements on RSs, the DSA aims to empower users through transparency and control, by allowing them to influence the socio-technical infrastructure by which they have been influenced for more than two decades. However, the way

in which such requirements have been implemented until now does not trace a consistent and substantive realization of the principle of user empowerment. The audits of Instagram and TikTok are emblematic of how the lack of guidelines for the implementation of art. 27 and 38 DSA can lead to disparate outcomes regarding which information, design features and control options are considered compliant under the DSA. These considerations allow us to come back to my first research question.

Addressing RQ1. *I argue that there is currently no unified interpretation nor consistent application of what the DSA provisions on RSs entail for VLOPs and, ultimately, user experience. However, related regulations such as CCD and EMFA can help clarify the path forward in terms of compliance.*

The first problem in this regard is the lack of definition of "parameters", "criteria" and "options", leading to a lexical vacuum which is often filled by the benchmarking convenience of auditors and platforms. The second problem is the current absence of legal guidance on what meaningful user control would consist in within this regulatory context: again, while the definition of thresholds is left to VLOPs, the benchmarking process, shared with auditors, may eventually capture the development of standards that should pertain to regulators. While my analysis has been based on three VLOPs, its results can be strengthened by considering the documentation pertaining to other VLOPs and VLOSEs and including also the (currently scarce) evidence on the DSA enforcement process by the Commission. As a result of the wider picture traced by the intersections of the DSA with related EU digital regulations, there is the normative ground to advance a meaningful personalization of online experience, intended as the ability for users to choose not only which type of content they want to see but also how they want to be exposed to such content. Meaningful personalization can be substantiated by allowing users, on the one side, to intervene on the design and logic of RSs through proportional and granular control options, and, on the other side, to select the degree of authoritativeness of the information they engage with and the diversity of the modalities through which they encounter such information. These perspectives on how to shape user

empowerment could be applied through an actionable framework for standardizing compliance.

Addressing RQ2. *Users can be empowered to achieve a meaningful personalization of their online experience through the operationalization of proportionality, granularity, diversity and authoritativeness. These perspectives contribute to a design framework for transparent and controllable RSs that may inform compliance standards for platforms under the DSA.*

The limitations of this contribution mainly concern its theoretical nature, which calls for empirical development and testing. In particular, further research should verify the effectiveness of my proposed approach through real-world experiments on user control of social media RSs, also by taking advantage of the forthcoming data access mechanism enabled by art. 40(4) DSA, which is now fully operational following the Commission's approval of the related delegated act (European Commission, 2025e). Directions worth exploring include: the empirical evaluation of the impact of personalization controls on user engagement, satisfaction, and perceived agency; the implications of feature-level contestability for interface design and users' awareness and trust; how to support exploration without sacrificing relevance or introducing cognitive overload; how to integrate authoritativeness signals into ranking processes to improve informed user choice, particularly in sensitive domains like news or health. Research efforts should not only aim to evaluate compliance, but also to redefine user empowerment through a meaningful personalization approach that views RSs as controllable and transparent tools to improve user awareness and self-determination in (very large) online platforms.

Chapter 5

A Critical Outlook on the DSA Enforcement Process

This chapter is dedicated to unfolding the functioning of the DSA enforcement process within the European Commission. It first provides an empirical analysis of the phases of an investigation against a VLOP through a case study of the proceedings against X. The evidence emerging from this analysis motivates a critique of the opacity of the Commission's decisions leading to the initiation of proceedings. The risks and potential consequences of this approach are exemplified through a theoretical scenario based on shadow-banning through RSs. I then propose a constructive approach to the evaluation of systemic risks for both compliance assessment and researchers' data access.

5.1 Introduction

As explained in Chapter 3, the EU legal provisions on which the regulation of RSs is based are limited in number and quite vague in scope. Their application is informed by compliance practices and enforcement actions, whose responsibility lies, for the former, on online platforms and, for the latter, on the European Commission and the national Digital Services Coordinators (DSCs). The analysis developed in Chapter 4 shows that large providers' compliance is not proactive, tending towards

minimal efforts or even transparency washing. The compliance practices of the large providers will arguably represent a model for the smaller ones: therefore, bad examples are likely to spread if not sufficiently challenged. Given that the Commission holds exclusive enforcement powers with regard to VLOPSEs, the directions and priorities of enforcement efforts against these providers are ultimately going to represent a blueprint for the application of the DSA requirements on RSs. Moreover, as occurred with the guidelines on the protection of minors based on art. 28, also the provisions of art. 27 and 38 could be detailed through guidelines or standards (art. 44(i)).

It is therefore useful to examine how the DSA enforcement process unfolds in practice to foresee how its shortcomings can impact the current and future implementation of the EU legal requirements on RSs. This chapter attempts to do so through a case study of the Commission's earliest enforcement action, i.e. the proceedings against X, which, despite an announcement of preliminary findings in 2024, have not been concluded yet. Subsequently, the problems raised by the opacity of the enforcement process are examined with regard to their political, procedural and normative implications, which are exemplified through a hypothetical scenario about the systemic risks enabled by RSs. This set of problems, as argued towards the end of this chapter, could be addressed through a pluralistic and transparent approach to enforcement, based on knowledge sharing and systematization.

5.2 A Case Study of Platform Governance under the DSA: the Proceedings Against X

The Commission has the exclusive supervisory and enforcement power for the obligations concerning VLOPSEs. To verify any failure to comply with the DSA, the Commission sent requests for information (RFIs) to various VLOPSEs, including X, Facebook, Instagram, AliExpress, TikTok and YouTube, starting from October 2023 (European Commission, 2024k). The RFI is the first stage of an investigation under the DSA, which can be followed by access to the data and algorithms used by the

platform, the conduction of interviews with informed subjects and inspections at the platform's premises. The information gathered through RFIs has already motivated the opening of various proceedings of infringement, including against X, TikTok, AliExpress, Facebook, Instagram and Temu between December 2023 and December 2024.

In the perspective of clarifying the structure and implications of the DSA enforcement process, I focus on the progression from an RFI to the initiation of proceedings against a VLOP: in particular, this contribution aims to provide insights on which type of evidence collected by the Commission through investigations can justify the decision to open a procedure of infringement under the DSA. Given that the text of already submitted RFIs is not publicly available, nor is it possible to access official documents about further investigative steps (as they may not have taken place yet), I will compare the press release about a specific RFI with its first legal consequence, i.e., the Commission decision on initiating proceedings against X (hereafter Commission decision), published on 18 December 2023 (European Commission, 2023a). Preceded by considerations on the problems raised by an opaque public governance of online platforms, this comparison will set the ground for understanding how the evidence collected by the Commission can justify the initiation of proceedings. This analysis relies on documents available through the European Commission website, including press releases, executive summaries, calls for applications and the Commission decision itself, which is the first legal document publicly disclosed regarding the DSA enforcement process, apart from designation decisions. The other decisions on initiating proceedings integrally publicly shared to date concern the one against AliExpress, the second of the three against TikTok (regarding TikTok Lite's "Task and Reward Programme", which has been suspended in the EU) and the one against Temu, in order of appearance (European Commission, 2024k). X is the only platform whose proceedings led to preliminary findings, which have been announced by the Commission through a press release (European Commission, 2024e).

5.2.1 The Problems Raised by Opacity

Notifying the initiation and development of proceedings against VLOPs through press releases, without publicly available legal documents backing most of the announcements, underscores the opacity of the communication strategy adopted by the Commission regarding DSA enforcement. The reason for sharing only a minority of the legal documents about the proceedings in such an inconsistent fashion has not been provided. However, the public availability of legal and policy evidence about the DSA enforcement procedures is essential for fostering scrutiny by researchers and platform users (Liesenfeld, 2025). Moreover, sharing information on who does what and how, i.e., which entity of the Commission is responsible for selecting specific evidence about the illegal activities of platforms and how such evidence can justify the opening of proceedings, is a necessary requisite for an accountable governance.

The lack of integral legal documents on the DSA enforcement strategy and on how the Commission evaluates DSA-mandated audits and information provided by VLOPs motivates concerns about platforms' compliance practices, outlining the risk of transparency washing (Zalnieriute, 2021) and regulatory capture. The need for transparency regarding the DSA enforcement process the has also been highlighted by an EU institution holding the Commission accountable: in fact, on 14 November 2024, the European Ombudsman (2024) issued a letter to the President of the Commission challenging the "general presumption of confidentiality" which, the Commission argues, applies to all documents related to DSA enforcement. In particular, following the Commission's rejection of a request to publish X's systemic risk assessment report in 2023 (which would have anyway been published by November 2024 as mandated by the DSA), the Ombudsman argues that "the requirement for professional secrecy laid down in Article 84 of the DSA [...] does not mean that all information the Commission obtains under the DSA should, as such, be considered as confidential" (European Ombudsman, 2024). In the Ombudsman's view, "applying a general presumption of confidentiality to the risk assessment report at issue [...] constituted maladministration".

In this regard, clarifications on the boundaries of what the Commission considers confidential information in the scope of the DSA are urgently needed to prevent accusations of maladministration, which might weaken public trust (Field and Roberts, 2020) and limit the overall effectiveness of the enforcement process. Indeed, the sharing of knowledge among the different social stakeholders involved, which would help design solutions for a pluralistic platform governance, is prevented by behind-closed-door enforcement and compliance dynamics. While the legal premises of the DSA align, in theory, with the principle of cooperative responsibility outlined by Helberger, Pierson, and Poell (2018), the practices of its application have until now followed a different approach, focused on the political bargaining between regulators and regulated entities. For example, while recital 90 DSA invites VLOPs to consult users' representatives for the assessment and mitigation of systemic risks, it is not possible to know, in many cases even after reading platforms' systemic reports, which organisations have been consulted; the same holds for the investigations carried by the Commission.

The Commission seems to interpret quite extensively the exceptions listed in art. 4 Regulation 1049/2001 (European Parliament and Council, 2001) regarding public access to its documents: indeed, "the protection of: — commercial interests of a natural or legal person, including intellectual property, — court proceedings and legal advice, — the purpose of inspections, investigations and audits" constitutes a sufficient ground to reject an access request, "unless there is an overriding public interest in disclosure". However, the Commission has not publicly stated the specific reasons for which DSA-related access requests have been or would be rejected following art. 4: its strategy may well be to reject all such requests until public interest in disclosure is established by the CJEU or national courts. While some information obtained by the Commission is certainly within "the purpose of inspections, investigations and audits", the fact that every non-public DSA-related document should be kept confidential to protect commercial, legal or investigative interests is an argumentative stretch, especially as regards evidence that can be obtained or generated independently from the DSA enforcement process.

5.2.2 The Content and Consequences of a Request for Information

According to the related press release, the RFI sent to X (European Commission, 2023d) on 12 October 2023 concerned "the assessment and mitigation of risks related to the dissemination of illegal content, disinformation, gender-based violence, and any negative effects on the exercise of fundamental rights, rights of the child, public security and mental well-being". The Commission aimed at scrutinising X's "policies and actions regarding notices on illegal content, complaint handling, risk assessment and measures to mitigate the risks identified". Following the submission of the RFI, the Commission decision identified five areas of concern in which the platform is suspected to have infringed the DSA provisions. A closer look at the document of the decision can help highlight the content and consequences of an RFI to a VLOP.

Firstly, Twitter International Unlimited Company (TIUC), "the main establishment of the provider of X in the European Union", failed to "diligently assess certain systemic risks in the European Union stemming from the design and functioning of X and its related systems, including algorithmic systems, or from the use made of their services" (European Commission, 2023a). In particular, the company did not "put in place reasonable, proportionate and effective mitigation measures" for "the actual and foreseeable negative effects on civic discourse and electoral process stemming from the design and functioning of X in the European Union": in fact, the current solutions "appear inadequate [...] notably in the absence of well-defined and objectively verifiable performance metrics" (ibidem). This failure is particularly evident in the moderation of content featuring languages different from English or pertaining to specific local and regional contexts. Suspicions that "insufficient resources [are] dedicated to mitigation measures" (ibidem) focus on the role of Community Notes[1], the collaborative feature that allows users to "leave

[1]In January 2025, Meta announced its plan to "end the current third party fact checking program in the United States and instead begin moving to a Community Notes program" inspired by X (Meta, 2025). This approach could not be adopted in the EU context thanks to the DSA: in fact, X's Community notes system has been challenged by proceed-

notes on any post and, if enough contributors from different points of view rate that note as helpful, the note will be publicly shown on a post" (X Help Center, n.d.).

Secondly, the company would systematically fail to process efficiently, "take decisions in a diligent, non-arbitrary, and objective manner" and answer "without undue delay" to the "notices [...] of allegedly illegal content hosted on X" (European Commission, 2023a); therefore, X's content moderation would be insufficient also when the input comes from users. Thirdly, the recent possibility of purchasing the blue checkmark that once marked an account as verified is considered deceptive and manipulative for users of X, who "are led to interpret [...] checkmarks as an indication that they are interacting with an account whose identity has been verified or is otherwise more trustworthy, when in fact no such verification or confirmation of trustworthiness appears to have taken place" (ibidem). Fourthly, the company did not abide by the transparency requirements for online advertising, "by not providing searchable and reliable tools that allow multi-criteria queries and application programming interfaces to obtain all the information on such advertisements as required by Article 39(2)" of the DSA (ibidem). Lastly, the VLOP seems "to have denied access to data that are publicly accessible on X's online interface to qualified researchers" (ibidem) by imposing costs for using the API and prohibiting the scraping of publicly accessible data, in violation of Article 40(12) of the DSA.

As the excerpts from the Commission decision highlight, the proceedings against X were initiated as a follow-up to the platform's response to the RFI submitted on 12 October 2023: in particular, the problematic issues regarding the handling of notices of illegal content, the measures to mitigate systemic risks and the platform's moderation policy, indicated in the press release about the RFI, correspond to the main areas of concern addressed by the decision. The consequentiality between the RFI

---

ings initiated by the Commission in December 2023 (European Commission, 2024k). Moreover, Meta signed the Code of Conduct on Disinformation (European Commission, 2025h), which represents a benchmark for compliance with related DSA provisions and, in its Commitment 31, explicitly requires signatories to "consistently use fact-checkers' work in their platforms' services, processes, and contents".

and the decision is underlined also by points 4 and 5 of the latter: in fact, while point 4 refers to the submission of the RFI and the response provided by X, point 5 states that, according to Article 66(1) of the DSA, "the Commission may initiate proceedings in view of the possible adoption of decisions [...] in respect of the relevant conduct by the provider of the very large online platform or of the very large online search engine that the Commission suspects of having infringed any of the provisions" (ibidem).

5.2.3 Next Steps of the Enforcement Process

The areas of concern addressed by the Commission decision include issues that emerge both from the bottom-up interaction with the platform and the top-down governance of its socio-technical ecosystem: data access by researchers and users' notices of illegal content pertain to the former, while the identification of systemic risks and the measures adopted to mitigate them pertain to the latter. The co-existence of top-down and bottom-up perspectives is not only motivated by the necessity to address issues coming from different sources (e.g., complaints filed by recipients of the service versus a letter from the Commission to the platform requesting clarifications); it is also functional to investigate different aspects of the same area of concern (e.g., the use of Community Notes to support content moderation, whose effectiveness is questioned both by users and by the Commission as the platform's reliance on such collaborative feature may hint at the lack of dedicated personnel for this delicate role). The structure and content of the decision indicate that, after the enforcement of the DSA, VLOPs and VLOSEs may not be able to consider anymore prima facie compliance as separated from the obligations to provide reasonably prompt and reliable answers to the issues raised by individual users and researchers. In practice, platforms are expected to go beyond the understanding of compliance as a checking-the-box effort; instead, they should engage in a substantive dialogue with users regarding content moderation and with researchers about data access.

The investigation initiated by the Commission decision aims to estab-

lish whether the failures outlined above "would constitute infringements of Articles 34(1), 34(2) and 35(1)" of the DSA as regards the first area of concern (inadequate assessment and mitigation of systemic risks), Article 16(5) and 16(6) for what concerns the second one (handling of notices of illegal content), Article 25(1) with respect to the third one (deceptive design of checkmarks), Article 39 with reference to the fourth one (lack of tools to ensure transparency in ads) and Article 40 apropos the fifth one (denied data access to researchers) (European Commission, 2023a). According to the European Commission (2023c), the next steps of the investigation will include gathering further "evidence, for example by sending additional requests for information, conducting interviews or inspections". Following the opening of formal infringement proceedings, the Commission will be empowered "to take further enforcement steps, such as interim measures, and non-compliance decisions" and "to accept any commitment made by X to remedy on the matters subject to the proceeding" (ibidem). Interestingly, the "DSA does not set any legal deadline" for the end of the proceedings, whose duration will depend on several "factors, including the complexity of the case, the extent to which the company concerned cooperates with the Commission and the exercise of the rights of defence" (European Commission, 2023c). The responsibility of carrying out the investigation pertains just to the Commission, whose decision "relieves Digital Services Coordinators, or any other competent authority of EU Member States, of their powers to supervise and enforce the DSA in relation to the suspected infringements of Articles 16(5), 16(6) and 25(1)" (ibidem). In this context, it is difficult to make hypotheses about a timeline for the conclusion of the proceedings.

The preliminary findings announced by the Commission in July 2024 (European Commission, 2024e) address three of the issues motivating the initiation of proceedings (dark patterns, advertising transparency and data access for researchers), while for the rest of them the investigation continues. The investigation was based on an "analysis of internal company documents, interviews with experts, as well as cooperation with national Digital Services Coordinators" (ibidem). As can be observed, neither the evidence sent by X in response to the RFI nor the content

of the interviews with experts was made available by the Commission, which has kept both the methodology and the results of its investigation opaque. The reported preliminary findings sent to X substantially correspond to the assumptions of the Commission decision, as the investigation outcome highlights that: firstly, "X designs and operates its [...] 'Blue checkmark' in a way that [...] deceives users", because, as "anyone can subscribe to obtain such a 'verified' status", users may be prevented from making "free and informed decisions about the authenticity of the accounts and the content they interact with"; secondly, "X does not comply with the required transparency on advertising, as it does not provide a searchable and reliable advertisement repository" allowing "for the required supervision and research into emerging risks brought about by the distribution of advertising online"; thirdly, "X fails to provide access to its public data to researchers in line with the conditions set out in the DSA"; for example, by preventing scraping or making the access to its API too burdensome or expensive (ibidem).

The Commission decision on sanctioning X, finalized in December 2025, was published in January 2026 by the United States House Judiciary Committee's Republicans as an alleged act of retaliation against the EU regulator (Jahangir and Hendrix, 2026). Its text confirms the scope of the aforementioned preliminary findings, while giving a sense of how the DSA enforcement team builds cases. From a methodological perspective, "the Commission relied not only on requests for information to X, whose responses are widely cited in the 183-page decision, but also on independent studies and interviews with experts and "its own evidence, as the alleged infringements were self-explanatory" (566)" (Fabbri, 2026). As conclusions "were reached through evidence-based resources mostly grounded in scientifically reproducible outcomes [...] researchers without access to confidential corporate information would, in principle, be able to reach the same results" (ibidem). The fact that reproducible scientific evidence was used to build the case against X highlights, once more, that the general presumption of non-disclosure for all the enforcement-related documents is not always justified by the confidentiality of the information handled.

5.2.4 The Intertwined Role of ECAT, DG Connect and External Contractors

Given the amount of investigative work that the Commission will embark on to enforce the DSA, it is useful to focus on the entities involved in this process: on the one side, the European Centre for Algorithmic Transparency (ECAT), within the Joint Research Centre (JRC), and, on the other side, the DSA enforcement team[2], within Directorate F (Platforms Policy and Enforcement) (European Commission, 2024h) of DG Connect. While the DSA enforcement team should have the responsibility of enforcing the regulation, some crucial investigative procedures, like "algorithmic system inspections" and "technical tests to enhance [...] the understanding of their functioning" (European Centre for Algorithmic Transparency, n.d.[a]), would be carried out by ECAT; therefore, it is unclear how the collaboration and the division of duties between the two institutional bodies will be managed. On the research side, the mission of ECAT is to study the "short, mid and long-term societal impact" of algorithmic systems, identify and measure "systemic risks associated with VLOPs and VLOSEs" and develop "practical methodologies towards fair, transparent and accountable algorithmic approaches, with a focus on recommender systems and information retrieval" (ibidem). On the enforcement side, the DSA enforcement team will be composed of "multi-disciplinary teams [...] co-operating with regulatory authorities in the Member States", which "will engage with stakeholders and gather knowledge and evidence to support the application of the DSA and to detect, investigate and analyse potential infringements of the DSA" (European Commission, 2024h).

The information provided by the Commission is currently not sufficient to understand how the work of ECAT and the DSA enforcement team intertwines specifically. Indeed, both these entities focus on gathering knowledge and evidence about VLOPs and VLOSEs to support the enforcement of the DSA, to the point that ECAT features an "Algo-

[2]Its latest denomination is Directorate Online Platforms: Society (CNECT.F): `https://op.europa.eu/en/web/who-is-who/organization/-/organization/CNECT/COM_CRF_244077`

rithm inspections and DSA enforcement" (European Centre for Algorithmic Transparency, n.d.[b]) team[3], whose work overlaps even nominally with that carried out at DG Connect. In particular, the profile of a technology specialist in the enforcement team at DG Connect is, if not similar, at least complementary to that of a researcher/inspector at ECAT. In fact, technology specialists "will work in seamless cooperation with the European Centre for Algorithmic Transparency (ECAT) and facilitate interactions with technical teams at very large online platforms and search engines" (European Commission, 2024j). The relevance of distinguishing between researchers at ECAT and technology specialists at the DSA enforcement team lies in the undefined borders of the presumption of confidentiality regarding the DSA enforcement (European Ombudsman, 2024). Although ECAT is a governmental research centre, no publication on the DSA, systemic risks and algorithmic systems inspections can be found in its publicly accessible repository (European Commission, n.d.). This underscores, again, an undeclared confidentiality policy that does not align with ECAT's aim of sharing "knowledge and facilitation of discussions on algorithmic transparency with international stakeholders" (European Centre for Algorithmic Transparency, n.d.). If the research on the DSA produced by ECAT is considered confidential, then it would fall under the scope of the DSA enforcement according to the presumption of confidentiality advanced by the Commission and challenged by the Ombudsman. If this were the case, however, there would be no distinction between the role of an ECAT researcher working on algorithmic inspections and that of a data scientist or technology specialist at the DSA enforcement team; therefore, it is not clear why these two overlapping roles would be included in two different directorates-general, i.e., the JRC and Connect.

Between November and December 2024, VLOPs and VLOSEs have published the systemic risk assessment reports describing the measures they undertook to comply with the DSA requirements on the identifica-

[3]Notably, the webpage corresponding to this reference, which detailed ECAT's teams (*Research* and *Algorithms & Inspections*) with the names of their respective members, has been recently removed or hidden, pointing to a voluntary reduction in transparency about the organisational structure of this research centre.

tion and mitigation systemic risks (art. 34 and 35), which lie at the core of the proceedings initiated by the Commission. However, the amount of work needed to verify the information provided by platforms cannot be carried out by the Commission alone, which will in fact rely on external contractors to produce the evidence needed to enforce the DSA systemic risk framework, as outlined in Chapter 2. In June and July 2024, respectively, two calls for tenders have been published by DG Connect to recruit external contractors to support the gathering of evidence for the DSA enforcement regarding VLOPSEs. The first call, which has an estimated total value 600.000 euros for a maximum duration of four months, involves carrying a study that "should inform with scientific and technical and other relevant insights the Commission's work on the first edition of a report on systemic risks and their mitigation, as referred to in Article 35(2) DSA" (European Commission, 2024a). Therefore, the research realized through the tender will inform the way in which the Board for Digital Services frames the definition, evaluation and mitigation of systemic risks under the DSA. The second call, which has an estimated total value of 12 million euros for a maximum duration of three years, is divided into three lots: the first one is aimed at developing "(real-time) monitoring and "early warning" systems and tools to inform and alert the European Commission about technological developments related to platform architectures, functions and features; economic and market trends; the emergence of systemic risks (including linked to the protection of minors, i.e., dangerous challenges on social media); as well as the emergence of potential digital threats"; the second one is devoted to "Evidence gathering and compliance monitoring of prominent and recurring risks in Member States and specific areas"; the third one concerns "online marketplaces and compliance with online advertising provisions" (European Commission, 2024b).

5.3 Challenging the Opacity of the Enforcement Process

The application of the DSA is proceeding at a fast pace: between the end of 2023 and the end of 2024, most social media VLOPs have been subject to investigations under the DSA, and four pornographic and two fashion retail platforms have been designated as VLOPs. While the DSA has come into force for every online platform since 17 February 2024, the main questions about the modalities and timeline of its application and the actions that platforms will need to take to ensure compliance are still open. In particular, it should be clarified which regulatory mechanisms and investigative evidence undergird the progression from an RFI to the initiation of proceedings against a platform. Considering that there is no legal deadline for the end of the proceedings, neither platforms nor users can have an estimate of their duration. Users are the main beneficiaries of the protection granted by the DSA, but if they are not involved in shaping its enforcement, the new provisions may not have tangible beneficial outcomes. The evidence on which the DSA enforcement relies risks being crafted by the interested interpretations of the regulatory requirements by auditing companies, while a wider stakeholder involvement does not appear under way. The previous case study shows the direction of the first proceedings of infringement under the DSA, thereby highlighting the type of evidence that European regulators see as symptomatic of non-compliance. The outcome of the ongoing proceedings against VLOPs will set a milestone for the future DSA enforcement strategies and their implications for users' online experience. The DSA has the potential to change the interaction between digital companies and EU citizens by enhancing public accountability and users' empowerment. However, for this potential to be realised, regulatory principles need to be translated into clear and transparent enforcement practices, which should be publicly disclosed to European researchers and citizens.

The public availability of integral legal documents about the DSA enforcement procedures is essential for fostering scrutiny by researchers and platform users. Moreover, sharing information on *who does what and*

how, i.e., which entity of the Commission is responsible for selecting specific evidence about the illegal activities of platforms and how such evidence can justify the opening of proceedings, is a necessary requisite for transparent and accountable governance. Therefore, the opacity of the enforcement process of the DSA, coupled with the absence of a definition of systemic risks, on which the Commission proceedings largely rely, may not be sustainable according to three arguments:

1. **Political**: the extent of the systemic risk sets, i.e., the collection of cases that will be considered relevant to a certain systemic risk by the regulator, depends on the power balancing and common interests between the regulator and the regulated entities, potentially making the implementation arbitrary. For instance, the change of the Commissioner responsible for DSA enforcement from Internal Market (Thierry Breton) to Tech Sovereignty, Security and Democracy (Henna Virkkunen) (von der Leyen, 2024) may imply a change of priorities for the enforcement.

2. **Procedural**: the procedures through which the Commission will fill the systemic risk sets with actual instances of platform design affordances, if they exist, are not public. The meeting minutes of the Board for Digital Services (European Commission, 2024i) do not mention any step in this regard. All the evidence needed should be collected by the DSA enforcement team in DG Connect, ECAT and tenders, but there is no publicly available evidence about the methods of such collection.

3. **Normative**: the above-mentioned risks of political arbitrariness and procedural opacity may weaken the methodological and normative solidity of the process through which systemic risks are recognised and evaluated, contributing to the chance that the Court of Justice of the European Union overturns the enforcement decisions made by the Commission.

5.3.1 A hypothetical scenario on the systemic risks enabled by recommender systems

The situation possibly arising from the intertwining of the three problems outlined above can be observed in the latest RFI sent to YouTube, Snapchat and TikTok on October 2nd 2024 regarding "the design and functioning of their recommender systems" (European Commission, 2024g). In particular, YouTube and Snapchat were asked "to provide detailed information on the parameters used by their algorithms to recommend content to users, as well as their role in amplifying certain systemic risks, including those related to the electoral process and civic discourse, users' mental well-being (e.g. addictive behaviour and content 'rabbit holes'), and the protection of minors". The questions concerned also the measures adopted by these platforms "to mitigate the potential influence of their recommender systems on the spread of illegal content". Further, TikTok was inquired about "the measures it adopted to avoid the manipulation of the service by malicious actors and to mitigate risks related to elections, pluralism of media, and civic discourse, which may be amplified by certain recommender systems".

As can be seen, the RFIs, whose content is not publicly disclosed and can be just inferred through the related press release, refers to various domains of systemic risks which may be impacted by the functioning of RSs. However, as systemic risks are neither defined nor subsume specific cases yet, it is not clear why the regulator chose to focus on the areas mentioned above for this RFI, nor how the areas of systemic risk considered in this RFI should be impacted specifically by the RSs of the selected VLOPs. It may be argued that the Commission is using this RFI to enhance its understanding of the impact of these platforms' RSs on the identified areas of systemic risks: if this were the case, the platforms, through their answers to the RFI, would be able to shape the enforcement process by exploiting the knowledge imbalance between the regulator and themselves. This argument outlines a procedural paradox, in which the regulator would rely on the knowledge of the regulated entities about their technology to enforce the regulation upon them. In this perspective,

a regulatory capture might emerge as a consequence of the lack of knowledge on how to include cases in each set of systemic risks. As a result, the balancing of interests between the regulator and the regulated would play a heavy weight in determining what should be considered compliant[4].

Moreover, the absence of guidelines or delegated acts about the implementation of DSA provisions on RSs (art. 27 and 38 DSA) hinders both the Commission and the platforms from tracing the boundaries of compliance regarding the design and functioning of these algorithms. As a consequence, the lack of accountability and transparency of the enforcement process may reinforce the outcomes that the DSA aims to prevent, at least in some cases. To discuss this perspective, I outline a hypothetical case-based scenario aimed at showing the practical problems that may emerge from the enforcement strategy currently adopted by DG Connect. The scenario is based on the following sequential assumptions:

1. Political priorities drive enforcement without transparent criteria.
2. Private negotiations replace public accountability.
3. Technical implementation obscures political choices.
4. Control by the Commission shapes digital spaces without democratic oversight.

Let us suppose that the Commission identifies protection of minors as a systemic risk in a short-video VLOP popular among minors, focusing on content that might affect their self-image and mental health. Indeed, European Commission President Ursula von der Leyen called for an "EU-wide inquiry on the broader impacts of social media on well-being" (Von Der Leyen, 2024, pag. 20) as a priority of her new mandate.

[4] A recent exchange on X between Elon Musk and Thierry Breton has highlighted the political debate influencing the enforcement process (Breton, 2024). Indeed, Musk's provocation about an "illegal secret deal" aimed at censoring "speech without telling anyone" emphasises the risks inherent to the bargaining aspect of the out-of-court enforcement process, whose development cannot be publicly inspected due to the absence of official legal documents about the methods and findings of the proceedings of infringement.

Thus, there is a pressing need of suitable design interventions in line with the DSA to limit the negative effects of RSs on users' mental health. In particular, in the case at hand, let us suppose that some LGBTQ+-tagged videos widely spread on this platform are found to include a distorted imagery about sexual minorities in the guise of self-help tutorials or psychological support. Given the quantity and reach of such content, which could not be removed through deletion due to its subtly harmful nature, which does not make it illegal per se, demotion is viewed as a potentially effective approach to content moderation as it limits the visibility of the content without outrightly removing it. Therefore, the Commission privately negotiates with the VLOP providers about how the RSs they deploy should treat LGBTQ+-tagged content. This negotiation faces competing political pressures, as the Commission needs to take into account the concerns raised by stakeholders with different political leanings.

Progressive groups argue that, while moderating content that is harmful for minors is a priority, the Commission should not limit youth's access to LGBTQ+ support resources. DG Connect should therefore require platforms to adopt an approach to content moderation based on a granular and widespread deletion aimed at removing the harmful content without *de facto* limiting, through demotion (Leerssen, 2023), the freedom of expression for well-intentioned LGBTQ+ creators. Conservative groups, on the contrary, argue that content deletion mechanisms, as they are often at least partially automated, are not flexible and precise enough to identify all and only the content that is harmful. Consequently, in the case of false positives (videos incorrectly identified as harmful), relying solely on deletion would imply an explicit limitation of freedom of expression, while demotion less so. The latter should therefore be preferred by the regulators, as it also has the advantage, in their perspective, of addressing an alleged over-exposure of minors to LGBTQ+ content that they believe may negatively influence their development.

Both groups agree on the necessity of reducing the exposure of minors to harmful content; their disagreement is on how to do so. It is at this point that their different political perspectives on freedom of expression and protection of minors come into play. Based on its political lean-

ing and the results of the negotiation with the platform, the Commission will decide which of the two courses of action to take, embracing the related consequences without previous public consultation and knowledge exchange. In both the cases, due to the lack of transparency in the DSA enforcement strategies, fundamental decisions about content visibility would occur through unaccountable executive action rather than democratic processes. In fact, it is difficult to prove that specific features of RSs are relevant to specific systemic risks (Reviglio and Fabbri, 2024). Further, the competing claims of different, or even conflicting, stakeholders problematize the Commission's attempt to include some algorithmic features among the enablers of such systemic risks.

Finally, the systemic risk assessment reports published by VLOPs at the end of 2024 underline that platforms currently choose individually which approach to use to deal with the challenges of scenarios like the one presented above. For instance, YouTube's report states that "Using recommender systems to order the presentation of content, including by elevating high-quality and trustworthy content, is a more proportionate approach to addressing harmful content risk than removing content altogether, which can present risks to freedom of expression and information" (Google, 2024b): therefore, this VLOP considers demotion as a fairer approach to content moderation than deletion. However, the need for consistent guidelines and coordination of compliance practices across platforms is put in evidence by the annulation of the Romanian presidential elections in December 2024, following which the European Commission opened the latest proceedings against TikTok focusing on the "risks linked to the coordinated inauthentic manipulation or automated exploitation of the service" through its RSs (European Commission, 2024c).

5.3.2 Evaluating compliance through a pluralistic approach

As the scenario presented above shows, there is fundamental normative dependence in systemic risk assessment, which cannot be enforced assuming an a-political standpoint. Every measurement framework for

platform harms embeds contestable normative judgments about what constitutes harm, which effects matter most, and how to balance competing values. To preserve value pluralism in DSA enforcement, the regulator should foster meaningful accountability through structured documentation requirements of systemic risk assessments, focused on methodology, engagement with alternative findings, clear metrics and thresholds, and regular evidence-based updates.

DG Connect and ECAT would play complementary roles in implementing this process. ECAT would systematically review documentation from multiple VLOPSEs in parallel to identify the contexts in which risk assessment methodologies might converge or diverge, recognise patterns and variations in how systemic risks are identified, and document cases where uncertainty about the conceptualization of such risks leads to different assessment approaches. Rather than evaluating cases in isolation, this cross-platform analysis, combined with a database of design alternatives and their associated risk levels, creates evidence-based accountability without mandating specific solutions. Building on this comprehensive analysis, DG Connect could expand this work by questioning platforms about threshold differences in identifying systemic risks where methodologies appear comparable, requiring justification for inconsistent results or when threshold choices lack a clear rationale, and acknowledging cases where uncertainty may motivate diverse approaches. This integrated division of responsibilities ensures that both technical analysis and regulatory oversight contribute to a more robust understanding of systemic risks while maintaining accountability through structured documentation. As VLOPSEs are already required to publish audit and systemic risk reports, the regulator should also make the results of its comparative analysis publicly available.

The normative nature of threshold decisions necessitates robust stakeholder consultation procedures, with the integration of external researchers' findings serving as a cornerstone of this process. This integration can operate at different levels, as research findings based on DSA-enabled data access may: (1) inform the development of risk assessment frameworks and metrics by the regulator; (2) help validate or challenge platforms'

risk assessments, indicating the direction for future investigations; and (3) contribute to identifying emerging systemic risks before they manifest at scale. This research integration process facilitates the public justification for starting proceedings of infringement, while ensuring decisions are grounded in empirical evidence. The goal is not to eliminate methodological differences, but rather to ensure that both platforms and regulators engage substantively with research-backed critiques and alternative approaches. This would create a dynamic feedback loop between research and regulation, where academic insights inform enforcement priorities and regulatory experiences guide future research directions. This approach to compliance and oversight shifts the focus from premature standardization to a dynamic process of structured comparison and evidence-based evaluation of systemic risks, requiring platforms (and regulators alike) to iteratively document their risk assessment methodologies, engage with competing research findings, and justify their choices[5].

5.4 Conclusion

No evidence collected through the investigations on cases under the DSA is, as of now, publicly available, and so is for most of the legal documents concerning the proceedings initiated against VLOPSEs. The text of the Requests For Information (RFIs) that the Commission sent to VLOPSEs under its supervisory power has not been shared. RFIs are the main instrument currently used by the Commission to gather knowledge on how platforms cause and address systemic risks: in fact, the investigative procedures available to the regulators include interviews to informed people, access to VLOPSEs' data and algorithms and even physical inspections at the platforms' premises, but it is unknown whether these procedures have been used by the DSA enforcement team, as they have not been announced through press releases like in the case of RFIs. Moreover, the legal texts of the Commission decisions on initiating proceedings against VLOPSEs have been publicly disclosed only in the case of

[5]The arguments outlined in this section were developed in collaboration with Michele Loi and Andrea Ferrario.

X, AliExpress and TikTok (only the one regarding the Lite Reward Programme, then suspended), while all the other ones have just been covered by press releases or summary notes (European Commission, 2024k).

The case study of the proceedings against X underlines the opacity that characterizes the public-facing side of the enforcement of EU platform regulation. The presumption of confidentiality that the Commission applies, in principle, to all the documents concerning DSA enforcement does not allow researchers and users to understand how and why enforcement priorities are decided. In fact, some *vertical* areas of intervention, like the protection of minors or the data access for researchers, are ostensibly prioritized both through the submission of RFIs and through the production of secondary regulation such as delegated acts and guidelines. Instead, other areas, such as the transparency and controllability of RSs, are not addressed either through standards (as indicated by art. 44(i) DSA) or secondary regulation: the related requirements are scattered across different legal texts addressing *vertical* areas such as disinformation and protection of minors. The requirements on interface design, explanations and user control of RSs are, however, the cornerstones of the user empowerment principle that the DSA and related regulations explicitly prioritize. Due to the opacity of DSA enforcement process, the motivations underlying this approach are not spelled out and cannot be reverse-engineered through the publicly accessible evidence. The hypothetical scenario on shadow banning presented above highlights the political, procedural and normative shortcomings of an enforcement approach based on opacity. To this end, only a transparent and pluralistic approach to the identification and assessment of infringements, based on a documented implementation of the systemic risk framework, could make the DSA enforcement process accountable.

The analysis developed in this chapter clarifies what trade-off between transparency and opacity is at stake in the DSA enforcement process, for whom it matters, and to what end. Institutional actors responsible for ensuring meaningful compliance with the DSA - primarily the Commission through DG Connect and ECAT, secondly Digital Services Coordinators - have the power of shaping enforcement priorities. Stake-

holders such as civil society organizations and independent researchers, who support the implementation of the DSA (Carvalho and Griffin, 2026), would benefit from access to the legal documents, methodological approaches and evidence base underpinning infringement proceedings and decisions — precisely what the Commission's general presumption of confidentiality currently withholds. I argue that this form of transparency is necessary for the accountability of the enforcement process, without which the assessment of systemic risks and the selection of enforcement priorities remain vulnerable to political arbitrariness and regulatory capture, as the case study of the proceedings against X illustrates. The dual-track approach proposed above — combining structured documentation requirements for platforms with researchers' data access under art. 40 — is therefore aimed at making the systemic risk framework legible and contestable for regulators, researchers and civil society alike.

Given the prominence of systemic risks enabled by RSs in the Commission's investigations, procedural transparency and contestability would also impact how the DSA requirements on RSs are understood and enforced. As highlighted in Chapter 4, the fact that standardization has not been pursued at all, despite being required by art. 44, is the consequence of the approach chosen by the enforcement leadership. Prabhat Agarwal, head of the DSA enforcement, explicitly acknowledged this shortcoming in his keynote at the DSA conference 2026: "a completely missing piece, as far as we can determine, is the standardization piece, which [...] no one had the courage to take it up [sic]"; relatedly, "the signal to noise in the audit reports is not what we wanted to be" (Agarwal, 2026). Had enforcement started from a the participatory standard-making process envisioned in art. 44, the landscape of audit and systemic risk reports would look quite different now. What emerges is therefore that platforms' minimal compliance with the DSA provisions on RSs is influenced by how the Commission interpreted and enforced them. However, in the absence of publicly available documents showcasing the Commission's approach in this regard, private contacts through policy networks become the main, if not the sole, way to collect relevant up-to-date information.

Chapter 6

Feeding the (short-video) feed: a design proposal for user control of social media recommender systems under the DSA

This chapter introduces a controllable and transparent RS for short videos integrated into an interactive user interface, through which a preliminary user study was conducted to provide insights into how the DSA-informed control features of RSs can enhance users' understanding and willingness to intervene on the recommendation process. After interacting with this platform, users participated in an evaluative survey featuring six dimensions: perceived control, perceived transparency, engagement with control features, perceived psychological workload, satisfaction, and impact on digital wellbeing. The findings indicate, among other things, that the availability of control features is associated with users' feeling of empowerment and perceived ability to recognize how recommendations steer their attention. These results open future research directions on the design of DSA-informed transparent and controllable RSs and provide practical guidelines for the implementation of the legal requirements for RSs.

6.1 Introduction

In the contemporary information age, RSs are widely used by online platforms to deliver personalized content to users, reducing information overload and enhancing their experience (Ricci et al., 2022). From scrolling through a series of short videos on TikTok to purchasing products recommended by personalized feeds on Amazon, users' choices and behaviour are continuously influenced by algorithmic recommendations. In particular, RSs influence how users access information and, indirectly, also how they think (Leerssen, 2020; Bonicalzi, De Caro, and Giovanola, 2023), exploit user emotions for targeted recommendations, and may eventually lead them to irrational decision-making (Grisse, 2023; Bartmann, 2023). Moreover, the black-box character of RSs often prevents users from understanding the rationale of recommendations and to contest inappropriate ones.

This lack of transparency threatens individual autonomy and raises ethical concerns (Milano, Taddeo, and Floridi, 2020). The Digital Services Act of the European Union is the first supranational regulation addressing the systemic risks resulting from RSs through provisions that, besides requiring online platforms to explain how their RSs work in the terms and conditions, empower users by allowing them to modify the parameters on which recommendations rely. Correspondingly, the major platforms have progressively provided explanations for their RSs and additional user control options (Meta, 2024b).

The effectiveness of the control features in practice depends mainly on their usability (C. Starke et al., 2025b; Bakalov et al., 2013). Whether or not the user actively intervenes on the recommendation logic is strongly influenced by the design of the UI (Busch, 2023), which in turn embodies the values and interests trade-off of the various stakeholders involved (Stray et al., 2023). Designing for controllable RSs is therefore a field still open to exploration.

To connect the previously presented evidence on VLOPs' compliance behaviour with the design of user control in RSs, this chapter introduces an interactive, explainable, and controllable short-video recommenda-

tion platform [1] [2] and undertakes a small-scale user study to evaluate the extent to which users' interaction with its RS is effective. The choice of short videos as the target domain of this research is motivated by the fact that short-video platforms are among the most popular social media services (Violot et al., 2024), significantly impacting the general public, especially young people. Overall, this contribution provides a preliminary insight into how the principles of proportionality, granularity and diversity in the user control of RSs, proposed in Chapter 4, can be implemented via a short-video interface. The outcomes of the user study carried out on this platform can motivate further research exploring the ethical and social impact of DSA-oriented RSs in online platforms.

Table 2 compares the main user feedback and control options of four VLOPs featuring short videos (Facebook, Instagram, YouTube, TikTok) and two short-video platforms popular in China (Douyin, BiliBili)[3], highlighting similarities and differences. While Western platforms mainly focus on feedback tools, their Chinese counterparts provide users with a wider set of options to intervene directly into the recommendation process (e.g. sliders for selecting the relative proportion of topics in the feed).

[1]The complete source code of this project is openly available at this link: `https://anonymous.4open.science/r/RecommenderOldApp-9E78/README.md`

[2]The frontend and backend of the platform were developed by two master students, Jingyi Jia and Pablo Jerez Arnau, as part of their thesis research at the Technical University of Munich (TUM) in 2023/2024. As the co-supervisor of these students' theses with Dr Wolfgang Woerndl, I contributed to design the interface and the topic-based RS on which the platform relies. The joint work resulted in the co-authored paper submitted to the International Journal of Human-Computer Studies on which this chapter is based, as stated in the Introduction (Chapter 1). This research project, including the short videos displayed on the platform, complies with TUM ethical procedures applicable to master theses, as it was developed as part of master theses at TUM and hosted on TUM server. These procedures do not require that the videos hosted on the platform be reviewed by an ethical committee.

[3]Chinese online platforms using RSs are subject to the Internet Information Service Algorithmic Recommendation Management (IISARM) Provisions (Reviglio and Santoni, 2023), the first ever regulation on RSs, since March 2022. This regulation adopts a different approach from the DSA, as it requires the platforms to be transparent and accountable to the central government, rather than to users, and follow the social values determined by the state.

Table 2: Overview of user feedback and control features across different short-video platforms. To facilitate comparison, the features have been grouped in five dimensions: *Feedback Entry Point* indicates how easily accessible the feedback and control options are to users (art. 27(3) DSA); *Dislike Reasons* include the motivations, if any, based on which users can give negative feedback to a piece of content; *Content Keyword Filters* are one of the most frequent control options across platforms, although with diverse affordances; the *Additional Control Options* complement the array of user control options compliant with art. 27 DSA; *Manage Feedback History* means that users can amend or withdraw the explicit data given by using the feedback and control options and that can be used personalized recommendations (art. 38 DSA).

Platform	**Feedback Entry Point**	**Dislike Reasons**	**Content Keyword Filters**	**Additional Control Options**	**Manage Feedback History**
Instagram	Click on "..." button followed by multiple navigations, or via settings	Creators, "uncomfortable", custom keywords and phrases	Select keywords and phrases associated with content to be excluded from the feed	Reset feed, select more, standard or less for political and sensitive content	Yes
YouTube Shorts	Long press on the video to display option menu	Descriptive adjectives such as "Irrelevant" and "Boring"	-	-	Yes
TikTok	Long press on video and select "Not Interested" for feedback, or via settings	Keywords, creators, background music, descriptive adjectives such as "Repetitive"	Select keywords associated with content to be included or excluded from the feed	Reset feed, activate STEM feed	Yes
Douyin	Long press on video and select "Not Interested" for feedback, or via settings	Keywords, creators, background music, descriptive adjectives such as "Repetitive"	Select keywords associated with content to be included or excluded from the feed	Slider to select the proportion of favorite topics in the feed	-
Bilibili	Long press for direct feedback, or via settings	Creators, descriptive adjectives such as "Irrelevant", channels and categories	-	Slider to select the proportion of favorite topics in the feed	-

6.2 Building a Short-video Recommender System: from Design to Implementation

To evaluate DSA-oriented user control and transparency features in a realistic setting, a recommendation-based short-video interface based on YouTube Shorts was implemented. The proposed interface includes functionalities that make the underlying RSs more granularly controllable by users. In this section, I discuss the dataset used, the interface design and the corresponding recommendation process.

6.2.1 Dataset

YouTube Shorts features videos up to 60 seconds in vertical format, optimized for mobile devices (Violot et al., 2024). Since YouTube does not specifically support API access for Shorts, a custom filtering technique was implemented to identify relevant videos. A script was developed to batch-download data and store it on a remote virtual machine using the YouTube Data API Java client. The keyword "short" was queried in the video textual features to find YouTube Shorts, as the video category for Shorts in the API includes only a minority of the total number of Shorts on the platform. The sampling pipeline used two YouTube APIs: the Search API to gather basic information about popular videos published within a specific time frame, and the Video API to retrieve detailed metadata such as video duration, hidden tags, and popularity statistics. Due to daily quota limitations, the sampling script was executed daily for a month, resulting in the collection of 206,743 videos.

However, not all these videos could be featured among candidate recommendations. The primary criterion for filtering the dataset was language. Since the user study was directed at EU users (the ones that are directly impacted by the DSA) the search was limited to videos in English and several European languages. The metadata provided by the videos contained significant gaps, particularly in language information. Consequently, it was necessary to predict the language of each video. For this purpose, a porting in Python of Google's language detection library

was used (Shuyo, n.d.). The textual features - title, description, tags - of each video were concatenated to predict its language. Subsequently, only videos in the following languages were retained: English, German, Spanish, French, and Italian. It should be noted that this method is not foolproof, as the language used in the textual features may not always correspond to the language of the video's audio content. This issue was observed in this interface, where several videos containing English only in the title, but not in the audio content, passed the filtering process. To achieve more robust language filtering, advanced methods leveraging the video's audio should be considered. A further manual filtering process was undertaken to refine the quality of the extracted video topics [4]. This process entailed evaluating the internal coherence of each topic and removing topics that lacked semantic consistency. Approximately 80 topics were filtered out of the initial 300, resulting in a final set of 220 topics. As a result of the whole filtering process, the original set of 206,743 videos was reduced to 39,296, corresponding to less than 20% of the initial volume.

6.2.2 Interface Design

When designing the interactive interface, particular consideration was given to ensuring that the system is transparent and accessible, striking a balance between providing user with suitable control features and keeping cognitive load acceptable. In fact, research in the music RSs domain found that, while "multi-level user control does not significantly increase cognitive load", personal characteristics of the user influence "recommendation acceptance and perceived diversity" (Jin, Tintarev, et al., 2020): therefore, the design of controls should be carefully assessed from a user-centric perspective. The interface design was based on the following guidelines:

- using visualizations, colours and dynamic feedback to assist comprehension;

[4]For a description of the topic modeling approach adopted, see subsection 3.3

- simplifying the user-system interaction;
- presenting control features with different levels of complexity in separate interfaces.

This approach allows the proposed RS to maintain transparency, reduce the extra cognitive load of user control features, and lessen the negative effects that stem from unfamiliarity with the recommendation process.

TikTok and YouTube Shorts were chosen as the primary design references for the platform, which recommends short videos retrieved from YouTube API. This decision is based on the idea that users would find a familiar interface easier to interact with. Users' familiarity with a platform often influences their understanding of the RS, thereby affecting their mode of interaction (Bart P. Knijnenburg, Reijmer, and Willemsen, 2011). People using social media applications or have prior computer science knowledge tend to use personalized RSs more effectively (Chernev, 2003; Aljukhadar, Senecal, and Daoust, 2012; Zhang and Chignell, 2001; Conati et al., 2014).

The design features of this short video recommendation platform may establish the basis for a standardization of the DSA requirements on RSs, which should cover "choice interfaces and presentation of information on the main parameters of different types of recommender systems, in accordance with Articles 27 and 38" (art. 44(i) DSA). Short-video VLOPs have been the target of various proceedings opened by the European Commission for suspected infringement of the systemic risk provisions of art. 34 and 35 DSA, especially in the areas of mental well-being and political influence and with regard to RSs. For example, Instagram is investigated for the design of its interface, "which may exploit the weaknesses and inexperience of minors and cause addictive behaviour, and/or reinforce so-called 'rabbit hole' effect" (Commission, 2024c). TikTok is subject to three proceedings, which concern, on the one side, its "persuasive design patterns", considered "problematic for the physical and mental well-being" of users (Commission, 2024d), and, on the other side, the role of its RSs in facilitating "coordinated inauthentic manipulation or automated exploitation of the service", supposedly linked to the recent

cancellation of the Romanian presidential elections (Commission, 2024b).

Compliance with the provisions on RS-enabled systemic risks is therefore one of the main areas in which these VLOPs are being held accountable under the DSA. Relatedly, clear design guidelines and standards are urgently needed for online platforms to adapt the control and feedback features of their RSs to the requirements of art. 27 and 38 DSA, as this will also affect how they plan to mitigate the systemic risks enabled by RSs. This contribution traces a direction in this regard, by proposing original algorithmic and interface design features that can be adopted by the main short-video platforms with minor refinement.

User Control

Following the classification framework for control levels proposed by (Jin, Tintarev, et al., 2020), the user control levels in this platform are categorized into low (providing explicit feedback on recommendations), medium (modifying user profile), and high (intervening on the recommendation algorithm). Figure 2 shows the interactive control features implemented and the level of control to which they belong.

- Low-level: Users can interact with the recommendations through an evaluative feedback mechanism, which includes signals (like/dislike, 5-star rating and reasons for their preferences) that the RS uses to compute the updated user preferences.

- Medium-level: Users can modify their profile by viewing and adjusting the personalized metrics the system calculates about the preferred topics and making secondary adjustments to the aggregated preference results.

- High-level: Users can directly intervene on the RS by modifying the number of videos provided by the two algorithms and adjusting the diversity of the recommendations in the hybrid RS at any time.

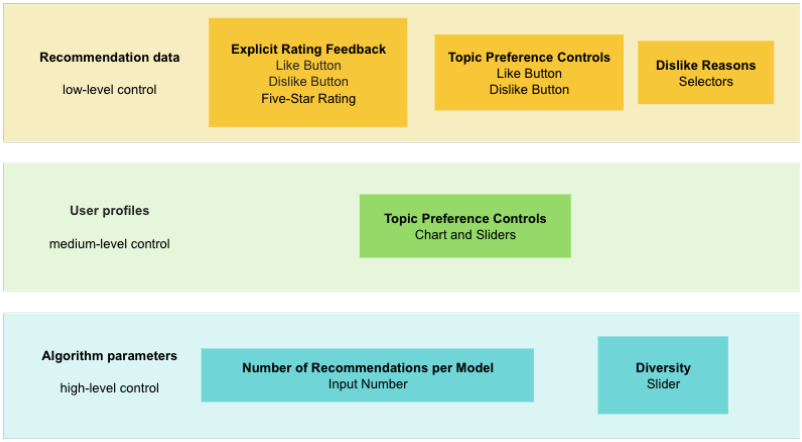


Figure 2: User control features in three levels

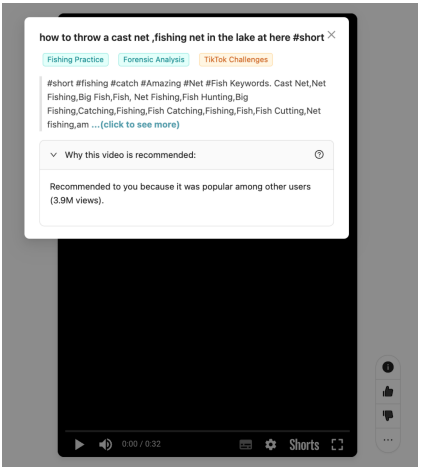
User Interface

The proposed interface consists of two main pages. Following the design strategy recommended by (Bart P. Knijnenburg, Reijmer, and Willemsen, 2011), simple and complex control features are presented in separate spaces. This design assumes that the video page will be used by all users and, therefore, should contain universally simple control features for broader customization. Expert users can switch to the user profile page for more advanced controls, depending on their interests.

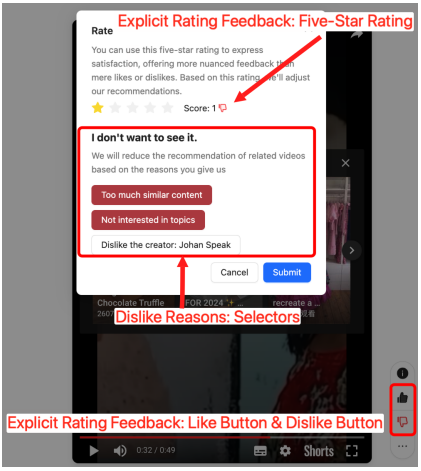
Dynamic feedback similar to the approach discussed by Bostandjiev, O'Donovan, and Höllerer (2012a) and Schaffer, Höllerer, and O'Donovan (2015) has been adopted to visualize the adjustment process, as it can enhance satisfaction and trust (Schaffer, Höllerer, and O'Donovan, 2015) and affect users' perceived accuracy in the recommendation results. Following the approach of Liang, Ponnada, et al. (2023) and Rani, Chu, and Mei (2022), pie charts are used to show the distribution of the user's preferences for video topics and tag clouds to describe such topics: in particular, according to Gedikli, Jannach, and Ge (2014), the use of word clouds for topic visualization can help users quickly grasp semantic con-

text while keeping cognitive load low. In addition, colours are used to convey meaning: green indicates positive factors that users want to enhance, while red indicates negative factors that they want to minimize.

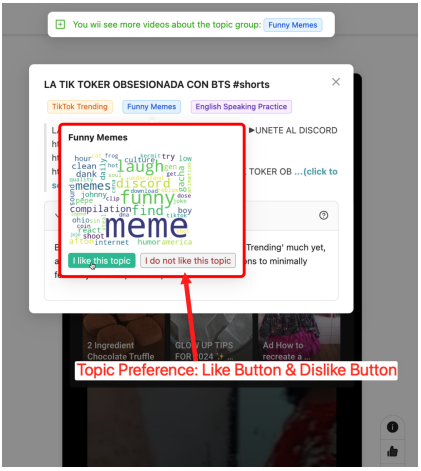
On the video page, illustrated in Figure 3 (a) - (d), users participate in the recommendation process through low-level control features, so they do not directly interfere with the recommendation process, but instead express their feedback allowing the RS to update accordingly. The user profile page (Figure 4) contains medium- and high-level controls. Each control feature is equipped with an information button that contains explanations of how the features work. The medium-level controls allow users to visualize their top-ten preferred topics in a pie chart calculated by the algorithm, and can modify the topic scores according to their preferences. For high-level control, users can intervene on the algorithmic parameters. As the platform's RS is hybrid, two algorithms (personalized topic-based and top-popular) provide candidate recommendations before a re-ranking algorithm generates the final list of videos. Users can adjust parameters to control diversity in the personalized topic-based algorithm, as well as the number of recommendations provided by each of the two algorithms. When the user hovers over the input number for the personalized algorithm, the slider associated with this algorithm is simultaneously outlined with a blue dashed line. This is done to visually emphasize their association. This interface design aims to operationalize the principles of proportionality and granularity introduced in Chapter 4 by distributing control across three levels (low, medium, high). This structure allows users to tune the extent of personalization in a proportional way and the level of granularity at which they intervene while keeping the cognitive load manageable. Diversity is addressed indirectly, through the exploitative/explorative slider for the personalized algorithm, which let users decide how much they want to be pushed towards novel topics versus familiar ones.



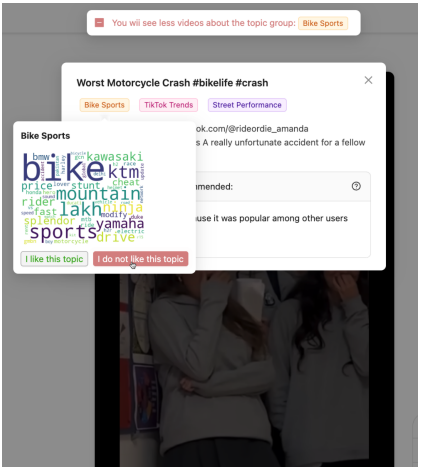
(a) Information button with topic tags and word cloud explanation.



(b) Extended feedback menu opened via the "…" button.



(c) Green prompt after clicking "I like this topic" under the word cloud.



(d) Red prompt after clicking "I do not like this topic" under the word cloud.

Figure 3: Feedback and control features on the video page.

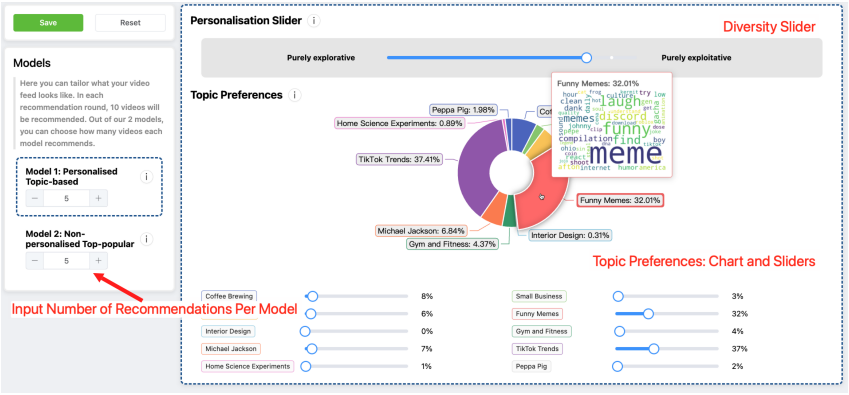


Figure 4: User profile page. Users can modify their topic preferences and recommendation diversity using sliders and control how many videos are recommended by each of the two algorithms using input numbers.

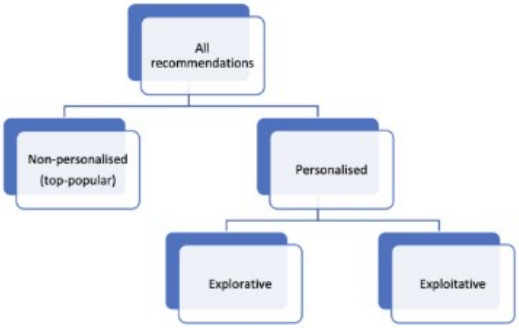


Figure 5: High-level recommendation logic. In each round, recommendations are categorized as either personalized or non-personalized. Personalized recommendations are further classified as either exploitative or explorative.

6.2.3 Recommendation Logic

In this project, a recommendation is conceptualized as a tuple consisting of a video and an accompanying explanation, which justifies why this video is recommended to the user. During each round of recommendations, the user is presented with ten video recommendation that can be either personalized or non-personalized, as showed in Figure 5. Non-personalized recommendations are based on video popularity, while personalized recommendations leverage user preferences. Personalized recommendations can further be classified as either exploitative or explorative. An exploitative recommendation aims to engage the user highly by satisfying their predicted preferences, e.g. by suggesting items similar to previously liked items. Conversely, an explorative recommendation aims to estimate the user's unknown preferences by suggesting content whose appeal to the user is not yet known, thereby enabling content discovery (McInerney et al., 2018). Through the control features in the user profile, users can specify the proportion of each type of recommendation they receive, while explanations for each recommendation are generated using a template-based approach explained below. The algorithmic implementation of control features and explanations is explained in the subsection *Recommendation Models*.

The system categorizes videos based on the *most relevant topic* in their topic distribution. Videos are recommended based on whether their most relevant topic aligns with the user's *most liked topics* or falls outside this set. The *most liked topics* are dynamically determined by the top ten topics within the user's preference vector. Each update to the preference vector triggers a corresponding update to the *most liked topics*, ensuring that the recommendations remain relevant and tailored to the user's evolving interests. The process used to update the user preference vector is described in the following subsection.

Modelling Video Topics

Recommender systems can be implemented through various techniques, such as collaborative filtering and content-based filtering. Given the ab-

sence of sufficient user-item interaction data in this dataset, which are necessary for implementing collaborative filtering, this system adopts a content-based filtering approach. This method utilizes detailed features of items to approximate a user's interest in those features, facilitating the grouping of similar items (Desrosiers and Karypis, 2011). The feature modelling technique should also allow for explainable and controllable recommendations.

The dataset, featuring short videos from YouTube, includes textual features such as video titles, descriptions, and tags, which are ideal for feature extraction via topic modelling. In topic modelling, a set of documents (in this case, the textual features of the videos) is used to extract a list of topics. Each document is then assigned a distribution over the topics, represented as a vector of topic scores. This means that each video has a topic score for each topic, which describes how relevant that topic is to the video, thereby allowing the grouping of videos using their most relevant topics, creating clusters with semantic similarity (Kherwa and Bansal, 2018). Similarly, users can also be modelled using a vector of topic scores, where the topic scores describe the user's interest in that topic. This approach fulfils the design goals of explainability and controllability, for the user is able to manually modify their topic scores.

Therefore, the choice of a topic-based model follows both the constraints of the available data and the design goals of the platform. The dataset provides rich item-level metadata (titles, descriptions and tags), but no longitudinal histories of user interactions, which makes collaborative filtering approaches inappropriate in this setting. At the same time, the system is intended to support explainability and user control: recommendations should be justifiable in terms of interpretable content features, and users should be able to adjust how much each type of content influences future suggestions. Modelling videos and users as distributions over semantic topics extracted from textual metadata satisfies both requirements: it allows the system to rely solely on item data, while exposing meaningful, human-readable features that can be referenced in explanations and manipulated through the control interface.

Various topic-modelling methods were considered before selecting

Non-negative Matrix Factorization (NMF) (Egger and Yu, 2022) for this project. Transformer-based approaches such as BERTopic can yield high-quality semantic clusters (Grootendorst, 2022), but they require substantially more computational resources and, in their standard form, assign each video to a single dominant topic. In contrast, the goal here is to obtain topic distributions that allow each video to belong to multiple topics at once, that can run efficiently in an experimental setting, and that can represent nuanced content associations. Among the alternatives that support such multi-topic representations, Latent Dirichlet Allocation (LDA) and NMF were compared. LDA builds on probabilistic assumptions and relies on raw word frequencies (Jelodar et al., 2018), but its non-deterministic nature makes repeated runs harder to reproduce consistently. NMF, instead, is deterministic and employs Term Frequency-Inverse Document Frequency (TF-IDF) to weigh the importance of words in a document, resulting in a weighted document-term matrix. These properties make NMF better suited to this system's need for reproducible, interpretable topics that can be directly linked to user-facing controls and explanations, without claiming any novelty at the algorithmic level.

The input documents for the topic modelling were constructed using the textual features of the data. The videos' title, description, and tags were concatenated and then textual processing was performed on them. This preprocessing involved several steps: lowercasing, lemmatization, removal of non-alphanumeric characters and links, and elimination of language-specific stop words. After preprocessing, each video was represented as a list of tokens, which served as input to the topic modelling algorithm. Topic modelling assigned a topic distribution to each video, represented as a vector of topic scores. This facilitates the understanding of complex semantic associations within videos. For some videos, a single topic score would dominate, indicating a strong association with that topic. In contrast, for other videos, multiple topics yielded high scores, suggesting semantic diversity and association with multiple topics.

After this basic text cleaning, an additional filtering step was applied to remove documents and tokens that were unlikely to contribute useful topic structure. Tokens that occurred in a very large proportion of videos

– for example extremely common hashtags such as “shorts”, “trending” or “viral” – were treated as non-expressive and discarded, as were tokens that appeared in only a handful of videos and thus provided little generalizable information. A grid search combined with manual inspection indicated that keeping only tokens that occur in at least 15 videos but in at most 10% of the corpus offered a good balance between coverage and expressiveness. Documents with fewer than 10 remaining tokens were also removed, as they contained too little text for reliable topic assignment. Applying these frequency-based and length-based filters substantially reduced the working dataset compared to the raw crawl, while retaining those videos whose textual metadata were informative enough for subsequent topic modelling.

Selecting the number of topics involves a trade-off between semantic coverage and interpretability. If too few topics are used, very different kinds of content are forced into the same cluster, making it harder for users to understand and act on topic labels; if too many topics are used, the model captures finer distinctions but produces a fragmented space that is difficult to present and control in the interface. To navigate this trade-off, several candidate configurations were explored through an inspection of the top terms for each topic. An initial solution with 300 topics was manually refined by examining the coherence of the resulting topic lists and merging or discarding those that lacked a clear semantic focus. The filtering process yielded a final set of 220 topics that still represented the diversity of the 39,296 videos corpus, in a way that could be meaningfully surfaced to users in explanations and control options.

Overall, the topic modelling component described above should be understood as the feature-engineering layer of a content-based RS rather than as a stand-alone original contribution to the literature. The NMF model transforms video metadata into a set of latent interpretable topics that act as item features, while the subsequent user modelling stage uses the same topic space to represent and update user preferences. In the remainder of this section, I describe how topic-based features and user vectors are combined to generate recommendations and how they support explainability and user control through the interface.

Modelling User Preferences

Modelling user preferences involved a similar method to modelling video topics, where each user's interest in various topics is represented as an explainable and controllable vector of topic scores that allows users to adjust their preferences. The vector of user preferences is updated during their interaction with the interface and leverages the user ratings to model the user's interest in each topic. Recommendations are generated in groups of 10. Once the user has watched all 10 videos, their interactions are used to update their vector of user preferences. There are two main ratings the user can give: video ratings and topic ratings. After processing the vector of topic preferences as described below, the vector is normalized.

A user can explicitly rate a video using a like or dislike, or use a more detailed 5-star scale, where 1 star equates a dislike, and 5 stars is a like. If the user does not provide an explicit rating, the watch completion ratio is interpreted as an implicit rating. Explicit and implicit video ratings update the vector of user preferences using an exponentially decaying weighted average (Figure 6). In this way, older ratings have a lower effect on user preferences (Ardagelou and Arampatzis, 2017), making recommendations more sensitive to newly expressed preferences.

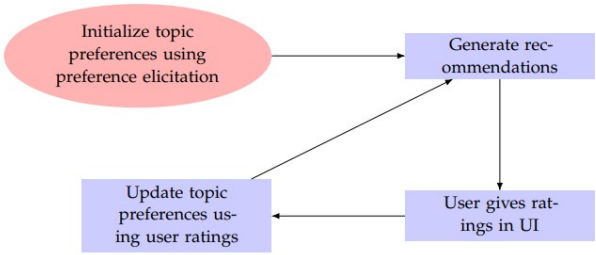


Figure 6: Recommendation-rating loop used to update the vector of user preferences

Equation 6.1 shows the formula used to update the vector of topic preferences using the video ratings and an exponentially decaying weighted

average: $T_{\text{user, new}}$ is the updated vector of topic preferences, $T_{\text{user, old}}$ is the old vector of topic preferences, T_{video} is the topic distribution of the rated video, $r_{\text{expl/impl}}$ is the rating, derived from the explicit or implicit ratings, using the tables 3 and 4.

The rating value $r_{\text{expl/impl}}$ depends on the given explicit or implicit rating. If the user gives an explicit rating, the system updates the topic preferences using the rating values in table 3. If the user has not given an explicit rating, the system uses the watch completion ratio as an implicit rating to estimate user's interest in the video. If the user watched more than half of the video, the topic preferences are updated using the formula 6.1 and the adjusted rating values from table 4. The user can explicitly rate an individual topic using a like or dislike. When a user likes a topic, the score of that topic increases. Conversely, when the user dislikes a topic, its score decreases. The formula used is shown in equation 6.2. For the top-ten preferred topics featured in the pie chart, their score (and prevalence) is visualised in real time through changes in the respective slider and section of the pie chart.

$$T_{\text{user, new}} = 0.9 \cdot T_{\text{user, old}} + 0.1 \cdot r_{\text{expl/impl}} \cdot T_{\text{video}} \tag{6.1}$$

$$T_{\text{user}}[t] = \begin{cases} T_{\text{user}}[t] + 0.1, & \text{if user likes topic } t \\ T_{\text{user}}[t] \cdot 0.5, & \text{if user dislikes topic } t \end{cases} \tag{6.2}$$

Table 3: Explicit ratings

user feedback (stars)	r_{expl}
1	-2
2	-1
3	0.5
4	1
5	2

Table 4: Implicit ratings

watch completion ratio	r_{impl}
0%–50%	0
50%–75%	0.5
75%–100%	1

Recommendation Models

Two distinct algorithmic models represent non-personalized and personalized recommendations. To enhance explainability and provide user control within the system, one model was developed for each recommendation paradigm. Users have the flexibility to choose the number of recommendations per model in each round, allowing them to influence the balance between personalized and non-personalized content. Below I detail the mechanisms each model employs to select videos and how the explanations accompanying each recommendation are tailored to reflect the underlying logic of the chosen model.

The top-popular model operates by randomly sampling one topic from a set. It then identifies videos where this topic is the most relevant and recommends the video with the highest view count from this subset. To maintain diversity in recommendations, the model tracks the topics it has sampled and avoids repeating them until all topics have been sampled at least once. After completing a cycle through all topics, the process starts anew in a second round. The explanation accompanying each video recommended by the top-popular model is tailored to reflect its popularity in the following way: ”Recommended to you because it was popular among other users (number of views)”. The view count is dynamically inserted based on the actual data associated with the video.

The personalized model categorizes its recommendations into exploitative and explorative. The proportion of each of these types is controlled by the user through a high-level control component, a slider. This slider ranges from ”pure exploration” (value 0) to ”pure exploitation” (value 1), adjusting in 0.1 increments. For example, if the user sets the slider

to 0.8 in a scenario where the personalized model is to generate 6 recommendations, this setting implies that 80% (5 recommendations) will be exploitative, while the remaining 20% (1 recommendation) will be explorative. Users' preferences are fundamental to determine the content of each type of recommendations. Exploitative recommendations feature only videos from the *most liked topics*: therefore, the higher the topic score, the more likely the topic is to be sampled. Explorative recommendations instead include both *rated topics* that are not among the most liked and *unrated topics*: this approach exposes the user to new topics, allowing them to discover new interests, while also revisiting topics that they have rated but are not among their *most liked*.

The personalized model's explanations consist of two parts. First, each recommendation states whether it is exploitative or explorative. For exploitative recommendations, the explanation reads "Because you seem to like topic T" indicating alignment with known preferences. For explorative recommendations, it states "Because you haven't explored topic T much yet" suggesting an opportunity for discovery. The second part of each explanation describes how the slider settings modified by the user influence the recommendations. This explanatory component changes dynamically based on the slider's position, providing tailored feedback to help users understand the impact of their settings on the recommendation outcomes. To provide more clarity to the user, the explanation adapts to whether the explained recommendation is explorative or exploitative. All the options for this second part can be seen in table 5.

6.3 Evaluation

A user study was conducted to evaluate the proposed approach to transparency and control. The study focuses on assessing whether the current design enables people with different levels of experience and domain knowledge to effectively use the RS while being aware of how they can influence the recommendations. It evaluates users' preferences for control features and investigates whether there are significant differences between low-, medium-, and high-level control. The study also explores

the extent of users' motivation to engage with the system and the reasons for their preference for specific control features. Focusing on users' subjective preference for specific control features is consistent with prior HCI and RSs research, which evaluates controllable interfaces primarily in terms of how much users interact with available settings and how satisfied they are with the resulting experience, rather than in terms of system-centric metrics such as predictive accuracy (Harper et al., 2015).

6.3.1 Participants

A total of 40 users were recruited to participate in the study through snowball sampling[5]. After excluding seven people due to anomalies in the interaction data (incorrect use of the device and non-compliance with the survey procedures as instructed), data from 33 participants was analyzed. The predominant age range was 20-30 years, with 32 participants in this age group and one outlier aged 57. The majority of participants were students and young working adults from Europe (10 from Spain, seven from Germany, two from Italy, and one from the Netherlands), with a further 13 participants from Asia (12 from China and one from Nepal). Figure 7 summarizes participants' self-reported frequency and type of interaction with social media platforms. The vast majority of participants showed a reasonable understanding of how RSs work and knew how to manipulate the system to deliver content they preferred, e.g. by giving positive ratings or by searching for and viewing such content more often.

6.3.2 Experimental Procedure

Participants were first given a brief introduction to the research background and objectives and invited to watch an introductory video walkthrough about the structure and the control and feedback components of the RS-based interface. Participants were then invited to fill out a short questionnaire about the frequency of social media use and to answer two questions about their *a priori* technical knowledge of RSs. Sub-

[5]Participants gave their informed consent to participate in the study.

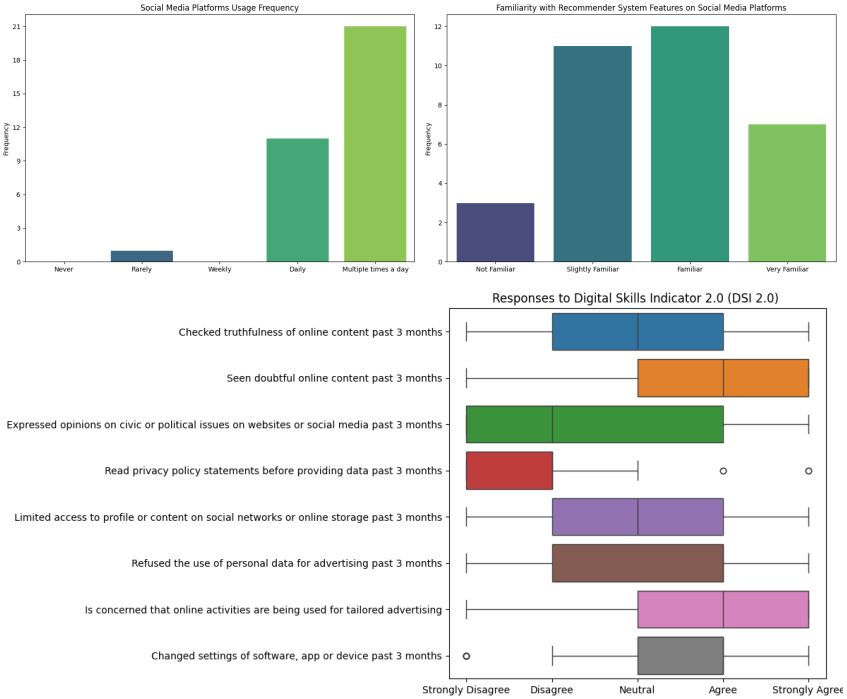


Figure 7: Social media experience. The two upper tables report the frequency of social media usage and familiarity with RS features; the lower table reports the responses to questions from the Digital Skills Indicator 2.0

sequently, during the preference elicitation phase designed to address the cold-start problem, participants were asked to select their preferred topics (between three and five) among a set of around 80 random tags.

The main interaction process, based on ten rounds of ten recommendations each, divided into three parts, started. In the first and last part, consisting of four rounds each, users scrolled through the recommended videos freely, simulating a real short-video recommendation platform. In the second part, consisting of two rounds, users received suggestions in the form of highlighted boxes that invited them to try various control features of the interface. By inserting explicit digital nudges between the two main sessions, the aim is to evaluate whether the extent to which users engage in active control of the recommendation process changes before and after they have been nudged to intervene on the parameters of recommendations. The system automatically collected interaction records during the interaction for subsequent data analysis.

At the end of the designated interaction time, users had to complete a survey to report on their experience. Following the multidimensional evaluation framework proposed by Tsai and Brusilovsky (2021), six critical aspects were considered to assess users' feedback and preferences about the control features of this short-video RS: perceived controllability; perceived transparency; engagement with control features; perceived psychological workload; satisfaction; digital wellbeing. The six dimensions were not connected a priori to specific interface features; rather, the evaluation framework was applied holistically to the interface in order to discover empirically which design choices contributed to which aspects of the user experience — an approach consistent with the exploratory scope of this study. As highlighted by Tables 6 and 7, the survey was mostly based on statements that users had to rate according to a five-point Likert scale, where 1 corresponded to *strongly disagree* and 5 to *strongly agree*. The Likert scale was especially suited to the dimensions related to the overall user experience from a specific perspective, such as *satisfaction* and *digital wellbeing*. Other dimensions, such as *perceived transparency, perceived controllability, engagement* and *perceived workload*, were aimed at gathering users' feedback on specific control features

from different angles and, for this reason, were not only based on Likert scale but also on multiple-choice or rank-order questions, asking them to order control features in a ranked list according to their preference. The quantitative approach of the survey was designed to allow the comparison of its results with the interaction data, which displayed users' actual use of the various control features: in this way, I could observe any inconsistency between users' stated preferences and effective behavior on the interface, to be then accounted for in the results.

6.4 Results

6.4.1 Perceived Controllability

Overall, the perceived level of control was relatively consistent across the different control features. Medium- and high-level control features generally provided slightly higher perceived controllability. Across all control features, *Topic Preference: Chart and Sliders* scored the highest despite being a medium-level control. This may be due to the use of a common visualization component, the pie chart, which directly displays the distribution of users' topic preferences, and the dynamic feedback effect, which shows the real-time impact of adjusting the sliders. The combination of these features allowed users to intuitively perceive changes in the RS under their control. This further validates the findings by (Bostandjiev, O'Donovan, and Höllerer, 2012a) and (Schaffer, Höllerer, and O'Donovan, 2015), according to which dynamic effects can provide a more significant sense of awareness and can further increase perceived control. 13 users provided more detailed feedback on the lack of control: 21% felt it was due to insufficient control options, 43% attributed it to poor video quality, and 28% felt it was due to the limited precision of the recommendation algorithm. The lack of accuracy of the algorithm and the quality of the videos might have influenced users' perceptions of control over the system as well, as some did not feel in control because, even after using the control features, the recommendations did not better match their taste. Allowing users not only to control the proportions

of displayed topics but also to add topics that are not displayed could enhance user control and satisfaction.

6.4.2 Perceived Transparency

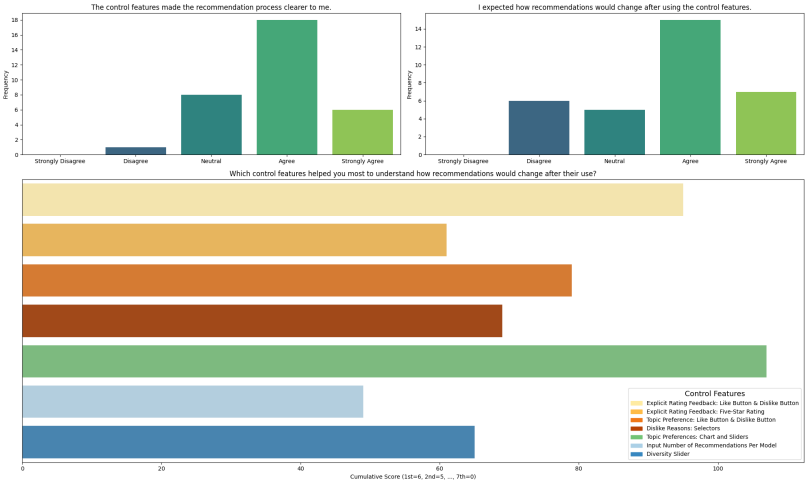


Figure 8: Perceived transparency. The two upper plots outline the perceived impact of control features on users' understanding of the recommendation process; the lower plot shows which feature users felt contributed most to their understanding.

The results for perceived transparency were generally positive: 24 users indicated that the control features helped them understand the recommendation process, aligning with the findings of Bart P. Knijnenburg, Bostandjiev, et al. (2012) that control features can enhance transparency. Most users could understand how the recommended content would change after using the control features, although seven users gave relatively negative ratings. Among all control features, the *Explicit Rating Feedback: Like button & Dislike button* next to the short video and the *Topic Preferences: Chart and Sliders* on the user profile page were the most effective in helping users understand how recommendations would change, with the *Topic Preferences: Chart and Sliders* receiving the highest rating.

The effectiveness of *Explicit Rating Feedback: Like button & Dislike button* may be due to its visual familiarity, as almost all online platforms have like and dislike buttons. The success of the *Topic Preferences: Chart and Sliders* could be attributed to the real-time dynamic feedback and the display of topic tags used by the recommender.

6.4.3 Engagement with Control Features

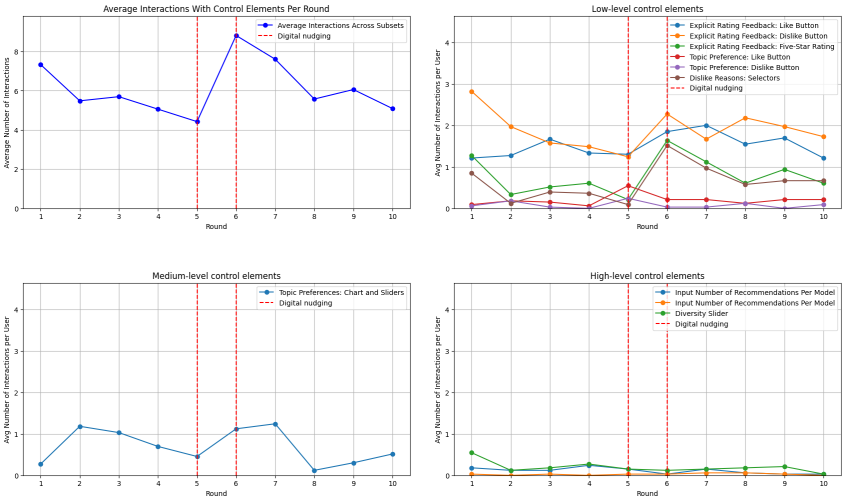


Figure 9: Comparative visualization of the average number of user interactions with control features of different level across multiple rounds of recommendations.

In this context, subjective engagement is intended as the extent to which users perceive to have interacted with the system, as observed in the survey results, while objective engagement corresponds to their actual level of engagement, measured through interaction data. Subjective and objective engagement with control features were compared using survey results and interaction history data, aiming to explore whether discrepancies exist between subjective evaluations and interaction data. In the survey, most users expressed their willingness to use the control

features of the RS (14 strongly agreed, 13 agreed, and four were neutral). Objectively, it was observed that users initially used the control features actively, but the frequency of interaction slightly decreased over time. Users might have been initially curious about the system, and thus tried different control features. Digital nudges increased the willingness to use control features in the two rounds in which they appeared, but in subsequent rounds the frequency of interaction gradually declined, aligning with the rounds without digital nudges (see Figure 9).

Among all components, *Explicit Rating Feedback: Like Button & Dislike Button* and *Topic Preferences: Chart and Sliders* ranked highest in terms of users' subjective preference for use, with the former perceived as the most intuitive and readily accessible feature, and the latter as easy to understand and providing direct and efficient control. Interestingly, *Topic Preferences: Chart and Sliders* was both the second most liked and the first least liked control feature in the related rank-order questions. This bifurcated evaluation applies similarly to groups with different levels of technical knowledge (Figure 10). Overall, users disliked medium- and high-level control features more, because of the complexity of the functionality and the possibly overabundant information. A minority of users were reluctant to use low-level control features, stating that such control methods were inefficient and they did not perceive timely changes in recommended content. However, the interaction data showed a different pattern compared to the survey results in some cases. Two components, *Explicit Rating Feedback: Five-Star Rating* and *Dislike Reasons: Selectors*, had low usage frequencies before digital nudges but increased ostensibly after the nudges. These two control features had low ratings in the survey, yet the interaction data showed that their use in later rounds was close to that of *Topic Preferences: Chart and Sliders*, which was the second most popular control feature after *Explicit Rating Feedback: Like Button & Dislike Button*, whose interaction count did not match with its high ratings.

To further evaluate the effectiveness of the control options provided, I performed a case-based analysis of the user study logs to match a user's interaction with the interface features and their recommendation output. In all the cases analyzed, I noted a direct impact of the users' choices

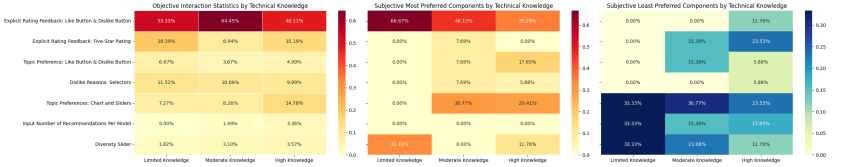


Figure 10: Heatmaps showing users' objective and subjective preferences for different control features

on the subsequent recommendations they received. I report here two emblematic examples representing opposite cases of the five-star rating feedback: one star, corresponding to *strongly dislike*, and five stars, corresponding to *strongly like*. In the first case, after the user gave one star to a non-personalized video recommendation associated with the topic *Mobile Gaming*, this topic was not featured in both the non-personalized and personalized recommendations of the next round. In the second case, a user attributed five stars to an explorative video recommendation associated with the topic *Music*; in the subsequent round, *Music* correctly appeared as one of the topics of the exploitative recommendations set.

6.4.4 Perceived Workload

User-centred UI design for RSs controllability can help reduce cognitive overload (McNee et al., 2003; Jin, Tintarev, et al., 2020). For this reason, the focus during the design phase was to minimize the perceived workload by using easily recognizable visualization techniques and colours to represent feedback. The assessment of perceived workload yielded positive results, with 48% of users stating that using control features did not require too much time and effort, and 60.6% indicating that understanding how to use the control features did not cause them undue stress or fatigue. Among all control options, *Topic Preferences: Chart and Sliders* in the user profile page caused the highest perceived workload, while *Explicit Rating Feedback: Like Button & Dislike Button* next to the video the least.

6.4.5 Satisfaction

Overall, the RS received positive evaluations, with most ratings for system satisfaction ranging from neutral to positive (21% neutral, 42% satisfied, 15% very satisfied). However, some respondents were not satisfied with their ability to control the recommendations (12% strongly disagree, 18% disagree, 12% neutral). It can be argued, similarly to what has been discussed with regard to perceived controllability, that this dissatisfaction may be attributed to the low quality of the videos and the lack of precision in the algorithm.

6.4.6 Digital Wellbeing

The survey results indicate a positive perceived impact of the RS on digital wellbeing (figure 11): the median response across all the four question was ”agree”. Most respondents indicated that the use of the interface enhanced their ability to avoid unwanted content[6] and gave them a sense of control over the recommendations, with the majority of responses falling within the ”neutral” to ”strongly agree” range. The effect of the system in increasing user exposure to preferred content may be limited by the video quality and the accuracy of the algorithm, as the most frequent responses ranged from ”neutral” to ”agree”. Most respondents also felt empowered to steer recommendations according to their aims and detected an improvement in their ability to recognize how RSs direct their attention (13 agreed, 13 strongly agreed).

6.5 Discussion

Based on the results and observations from the user study, insights have been derived that can guide the development of future controllable and transparent RSs.

[6] In the context of this study, unwanted content refers to videos or topics that were not aligned with users' preferences. No participant reported being exposed to content perceived as harmful or toxic.

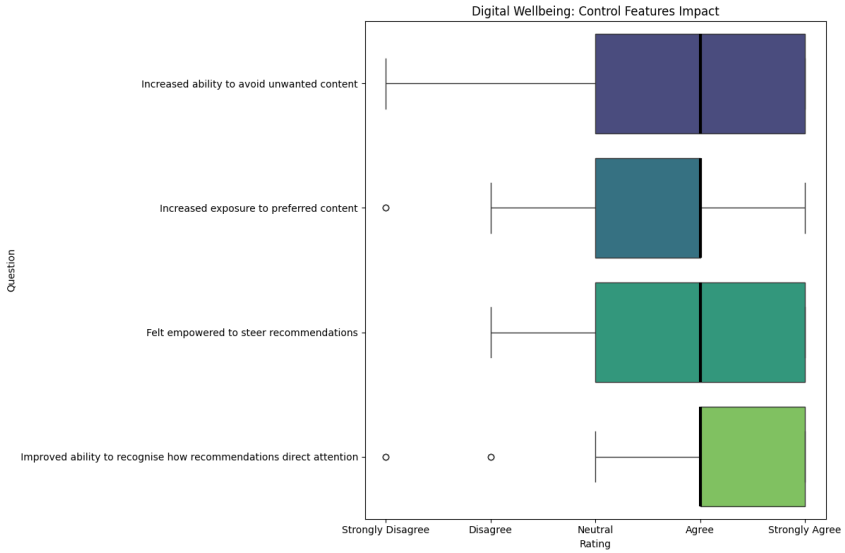


Figure 11: Digital Wellbeing

Users tend not to engage directly in the recommendation process. Although the majority of respondents expressed a preference for having control features, both subjective responses and objective interaction data demonstrate that they are weary of using these features. Indeed, providing many user control options may come into conflict with the original aim of any RS, i.e. reducing information overload.

Users with different levels of domain knowledge have similar preferences about the levels of control features. Figure 10 illustrates that the preferences for different control levels among users from different backgrounds converge. Users tend to prefer explicit feedback (like/dislike) more than medium- or high-level controls. Users with high domain knowledge have more dispersed preferences: their dislike for medium- or high-level control features is not as pronounced as for the other two groups.

Providing transparent and user-friendly control features can encourage their usage. The most used control features typically exhibit a

transparent and user-friendly design allowing users to easily access and use control features, which are also described through explanations. For instance, *Topic Preferences: Chart and Sliders* enable efficient adjustment of topic preferences and are appreciated for their ability to clearly delineate how topics are ranked by the algorithm.

The availability of control features is associated with users' feeling of empowerment and perceived ability to recognize how recommendations steer their attention. The levels of control reflect the depth of user engagement with the interface design and algorithmic structure of the RS. Regardless of the extent to which control options are used, their presence stimulates users' perceived empowerment and awareness on how RSs work and direct their attention.

These insights can help develop a meaningful approach to substantive compliance with art. 27 and 38 DSA, going beyond the limited explanations and control options currently provided by VLOPs. It can be argued that platforms do not have incentives, beyond regulation, to implement a transparent and controllable recommendation process, which may eventually hinder their business model oriented at optimizing user engagement, time spent and growth. Moreover, as the DSA does not detail what "parameters", "criteria" and "options" for users to steer RSs should consist in, platforms may tend to minimize their effort to provide granular control tools like those envisioned in this study. However, this argument can be challenged from two perspectives. First, user interactions with granular control and feedback tools would result in explicit signals about their preferences, which can, in turn, be used by platforms to provide more precise and useful recommendations (Fabbri, 2023c): hence, supporting users' self-determination can represent an economic incentive for platforms through the collection of explicit and, therefore, better quality data on users' preferences. Second, the recently adopted guidelines on the protection of minors under the DSA invite platforms to "prioritise 'explicit user-provided signals' to determine the content displayed and recommended to minors", to "explain why each specific piece of content was recommended to them", and to allow "them to select content categories and activities they are most or least interested in",

thereby specifying the transparency and control requirements of art. 27 (European Commission, 2025f); while not compulsory and just related to the protection of minors (art. 28), these guidelines may end up representing a compliance benchmark under the DSA and therefore constitute a further incentive, also directed at increasing users' trust. In sum, platforms have clear economic and regulatory incentives in the EU context to make RSs transparent and controllable.

This study traces an actionable perspective in this respect, as it features a proof-of-concept for user control that can be applied to large-scale environments without major issues. In fact, the scalability of topic modeling in real-world systems is well documented by existing deployments across major platforms, which use topic inference mechanisms in content moderation and recommendation tasks, among others. For instance, YouTube's content classification system has been used to create a public dataset with eight million labeled videos (Abu-El-Haija et al., 2016). Moreover, the absence of explicit textual metadata does not preclude topic modeling in large-scale environments: vision-language models such as CLIP (Radford et al., 2021) and multimodal transformers (Gabeur et al., 2020) enable unsupervised zero-shot classification and semantic indexing directly from visual and auditory cues. The extended use of RSs relying on uninterpretable embeddings or deep neural architectures (Covington, Adams, and Sargin, 2016) does not hinder controllability either, because user-facing controls can operate independently as a filtering layer for recommendations without requiring that the underlying ranking model be interpretable. In practice, when a user decides to avoid videos of a certain topic, the internal representation of those videos employed by the RS does not related to how the user can express their feedback on the interface; those videos would simply not be part of the filtering process. In fact, various user control approaches are already in use in short video VLOPs, such as *Change sensitive content* on Instagram (Meta, 2024b) and *Keyword filters* on TikTok (TikTok, n.d.). These examples highlight the technical feasibility of incorporating interpretable, topic-based user controls at scale, an approach that has traditionally been overlooked by platforms before the implementation of the DSA.

6.6 Limitations and Future Work

The limitations of this system are mainly due to the quality of video data and the relatively simple algorithmic model adopted, which does not include collaborative filtering. The topic-based NMF approach, while compatible with the controllable interface and updatable in real time, deliberately reduces the number of variables compared to commercial systems; this choice likely decreases both the accuracy and coverage of recommendations and favours videos with richer metadata, so findings about user experience should not be read as performance benchmarks for large-scale platforms. In addition, the implementation relies on textual features and does not exploit video-specific signals such as visual or audio cues, so the interface is not fully tailored to the specificities of short-video consumption. Furthermore, I acknowledge that the type of participants selected for the user study is skewed towards young individuals very familiar with online platforms. The study primarily targeted people aged 20-30 who exhibited a high frequency of social media use and had a relatively good understanding of RSs. An additional methodological limitation concerns the absence of granular baselines in the experimental design due to low resources. Since all participants had simultaneous access to the full interface — comprising low-, medium- and high-level control features — the study cannot isolate the contribution of any specific level of control granularity to the observed outcomes while collecting participants' perceptions. Further studies should take into account the preferences of people in different age groups with varied social media usage patterns and limited understanding of the technical underpinnings of online platforms and algorithmic recommendations. Future contributions are also expected to set a closer focus on the impact of transparency and control for short-video RSs on specific dimensions of user experience with high societal relevance, such as wellbeing. Developments could also be obtained through the access to platforms' non-public data granted to vetted researchers by art. 40(4) DSA starting from 2025, as the related delegated act was published in July 2025 (European Commission, 2025e).

6.7 Conclusion

This chapter focuses on the design of a DSA-informed transparent and controllable RS. An interactive recommendation platform for short videos was introduced and a user study was conducted to set the ground for further research exploring how user control features in RSs can be designed so that users can understand and are willing to intervene on the recommendation process. The interface's control features were categorized into three levels of complexity. Users' feedback on the system was discussed across six dimensions: perceived control, perceived transparency, engagement with control features, perceived psychological workload, satisfaction and digital wellbeing.

Taken together, the results provide a first, though limited, answer to each of the six evaluative dimensions. Perceived controllability and transparency are generally positive, suggesting that the control features help users understand and intervene in the recommendation process, even if some respondents attribute residual lack of control to video quality and algorithm accuracy rather than to the interface itself. Engagement shows that users are curious and willing to experiment with controls, but that sustained use declines over time, raising questions about how to design controls that remain attractive beyond initial exploration. Perceived workload is acceptable on average, yet medium- and high-level controls are perceived as more demanding, highlighting a tension between granularity and usability. Satisfaction is mixed: while many users are satisfied with the overall system, a non-negligible share are dissatisfied with their ability to control recommendations. For what concerns digital wellbeing, users feel more able to recognize how recommendations steer their attention than to reliably increase exposure to preferred content or avoid unwanted content.

This contribution traces a direction for the implementation of the DSA requirements on RSs, by proposing control features for RSs that could be adopted by online platforms with minor refinements to ensure substantial compliance with this regulation. Moreover, despite the fine-tuning needed to improve the user-friendliness of this simulative environment,

the implementation of the control features for short-video recommendations presented may have a beneficial impact on users' empowerment and awareness of how RSs direct their attention in online platforms. Since European Commission President Ursula von der Leyen called for an "EU-wide inquiry on the broader impacts of social media on well-being" (Von Der Leyen, 2024) as a priority of her new mandate, there is a pressing need of suitable design interventions in line with the DSA to limit the negative effects of RSs on users' mental health. This study is at the forefront in addressing the aforementioned gap in policy-oriented socio-technical research.

Slider value	Exploitative recommendation	Explorative recommendation
1 (pure exploitation)	you want your personalised recommendations to exclusively focus on your most liked topics.	-
0.8, 0.9	you want your personalised recommendations to heavily focus on your most liked topics.	you want your personalised recommendations to minimally feature your unexplored topics.
0.6, 0.7	you want your personalised recommendations to rather focus on your most liked topics.	you want your personalised recommendations to feature some of your unexplored topics.
0.5	you want your personalised recommendations to represent both your most liked and unexplored topics equally.	
0.3, 0.4	you want your personalised recommendations to feature some of your most liked topics.	you want your personalised recommendations to rather focus on your unexplored topics.
0.1, 0.2	you want your personalised recommendations to minimally feature your most liked topics.	you want your personalised recommendations to heavily focus on your unexplored topics.
0 (pure exploration)	-	you want your personalised recommendations to exclusively focus on unexplored topics.

Table 5: Second part of the explanation for personalized recommendations

Aspect	Questions
Perceived Controllability	1. For each control feature: Please rate the extent to which you feel this feature allows you to control the content the system recommends. 2. Optional: What do you think is the main reason for feeling a lack of control when using this recommender system? Please select the most applicable reason or specify in 'Other': [ ] Complex interface design. [ ] Insufficient customization options. [ ] Inadequate recommender algorithm or failure to update recommendations promptly. [ ] Video quality is not high. [ ] Other
Perceived Transparency	1. The control features made the recommendation process clearer to me. 2. I expected how recommendations would change after using the control features. 3. Which control features helped you most to understand how recommendations would change after their use? Please rank them from most to least helpful.
Engagement with Control Features	1. Please indicate which control feature you are most inclined to use. 2. Please indicate which control feature you are least inclined to use. 3. I am willing to use the user control features of the recommender.

Table 6: Survey questions for measuring users' perceptions – Part 1

Aspect	Questions
Perceived Workload	1. Adjusting the settings of the recommender system does not require a lot of time and effort. 2. Understanding how to use the user control features does not cause me stress or exhaustion. 3. Please indicate which control feature requires the most effort to be used and briefly explain why. 4. Please indicate which control feature requires the least effort to be used and briefly explain why.
Satisfaction	1. I am satisfied with my ability to control the recommended content. 2. I am satisfied with the overall experience of the recommender system.
Digital Wellbeing	1. The use of control features increased my ability to avoid content that I prefer not to see. 2. The use of control features increased my exposure to the content that I prefer. 3. The availability of control features made me feel empowered to steer the recommendations to content I find interesting, even if I didn't use the control features most of the time. 4. After interacting with this interface, my ability to recognize how recommendations direct my attention to certain categories of content has improved.

Table 7: Survey questions for measuring users' perceptions – Part 2

Chapter 7

Further User Perspectives on the Controllability of Recommender Systems

To expand the experimental evidence of the previous chapter, a further small-scale user study was run on an updated version of the short-video platform. The technical updates were aimed at improving user experience through topic relevance, accessibility of controls and ease in video scrolling [1] *The study aims to compare user feedback and interaction with the interface across two different sets of users: unguided, who completed the study autonomously, and guided, who received instructions and support by the researcher. While still at a prototypical stage, the observation of user behaviour and interventions sheds light on the importance of algorithmic literacy for an effective user empowerment.*

[1] The technical changes to the platform were implemented by Jilin Liao as part of the research for his master thesis at the Technical University of Munich (TUM), jointly supervised by Dr Wolfgang Woerndl and in 2024/2025: "DSA-Driven Research to Improve Usability and Accuracy of Recommendation Systems in Short Video Platforms". I curated the interface and UX design concepts, which constitute the basis for the technical implementation and the user study described in this chapter. All the code and documentation of the new version of the platform can be found at: `https://github.com/windsound3482/Recommenderapp`. The user study was funded by the Chair of Connected Mobility at TUM. This research project, including the short videos displayed on the platform, complies with TUM ethical guidelines applicable to master theses, as it was developed as part of master-level research at TUM and hosted on TUM server.

7.1 Introduction

Building on the findings of the previous chapter, the user study presented below is aimed at collecting additional evidence on how control options for short-video RSs inform users' perception of empowerment and understanding of the recommendation process. The short-video platform introduced in Chapter 6 was updated to enhance topic relevance, accessibility, granularity of controls and ease in scrolling through recommended videos. After describing the technical updates of the dataset, interface and algorithmic infrastructure of the platform, this chapter reports the rationale, design and results of the user study. Overall, the objective of this second study is to expand the dimensions related to perceived controllability, perceived transparency and digital wellbeing of the one presented in Chapter 6.

Compared to the first study, this chapter deepens the empirical analysis in three respects. First, it addresses some limitations of the original topic-modelling pipeline by improving topic relevance and coverage through an updated dataset that combines YouTube's native video category taxonomy with refined topic-based controls, while still operating within an interpretable, text-driven framework. Second, it shifts the focus of the evaluation towards social aspects that are central to the DSA debate, namely participants' algorithmic awareness, their perceived ability to steer exposure and diversity of content, and their assessment of the effectiveness of interface design for protecting minors and vulnerable users, as per the rationale of the Commission's guidelines on art. 28 DSA (Commission, 2025). Third, by contrasting unguided and researcher-guided user interaction with the platform, the study explicitly probes the role of algorithmic literacy as a precondition for meaningful control, building on prior evidence that users who better understand how feeds are curated feel more empowered to question and reshape recommendation outcomes (Eslami et al., 2015; Bart P Knijnenburg et al., 2012). Focusing on these dimensions in the platform design and final survey represent a way of deepening the application of the principles of proportionality, granularity and diversity outlined in Chapter 4.

7.2 Platform Updates

7.2.1 Dataset Improvement

Table 8: Retained video categories (with ID and name) from YouTube API.

Category ID	Category Name
1	Film & Animation
2	Auto & Vehicles
10	Music
15	Pets & Animals
17	Sports
19	Travel & Events
20	Gaming
22	People & Blogs
23	Comedy
24	Entertainment
25	New & Politics
26	How to & Style
27	Education
28	Science & Technology
31	Anime/Animation

In the YouTube Data API, Shorts are a considered a video category (whose ID is 42), like "Music" or "Gaming", rather than a resource. However, many short videos are not categorized by YouTube as Shorts, but instead placed in other video categories, which are 31 in total. For each of these categories, the 250 most viewed and liked videos were identified through the filters available in the search function and then cross-checked using the approach proposed by Violot et al. (2024) to verify whether they were Shorts. This approach consists in sending "a GET request to `www.youtube.com/shorts/<videoId>` for each videoId and checking[ing] in the redirection link if the URL stayed the same or if it was modified to the regular `www.youtube.com/watch?v=<videoId>`": this allows to verify whether YouTube classifies the retrieved video as Short. Only 15 categories were retained (Table 8), as they included a suf-

ficient number of short videos, ranging from 17 to 172. Using the video category ID 42 and the redirection approach allowed to overcome the video quality issues of the previous method based on filtering videos retrieved through the hashtag #short(s). Moreover, this approach discounted the need for topic modelling, which introduced the researchers' bias in the topic-based classification: in the new version of the platform, topic attribution and labels match exactly those embedded in YouTube API, thereby establishing a closer correspondence between the prototype and the real-world social media.

The use of the 250 most viewed and liked videos per category could raise the concern that highly popular content might also include toxic or otherwise inappropriate material [2]. Videos were sample-checked to exclude potentially illegal and policy-violating content that might have been disregarded by YouTube's content moderation process. It should also be noted that, within the study reported here, no participant indicated exposure to harmful or toxic content. When respondents referred to unwanted videos in the survey and qualitative feedback, they consistently used this term to describe content that did not match their interests or preferences, not content that they considered inappropriate or otherwise negatively impacting their wellbeing.

7.2.2 Interface Design

The platform interface has been modified with respect to the one presented in Chapter 6 to facilitate user engagement with the control features. On the main interface, the video, with the feedback bar on its right, appears in the centre of the screen, while the user control tabs[3] are displayed on the left in a collapsible format. The most noticeable change in the feedback bar (figure 12) is the separation between *like/dislike* (showed

[2] As highlighted above, the short-video platform and its dataset were developed as part of master's theses at the Technical University of Munich (TUM), which does not require a separate submission to a research ethics committee for the videos deployed on a platform developed for master-level work. Compliance with the university's ethical guidelines was verified in the supervision process.

[3] All interface features are accompanied by an information button that opens a tab explaining their functioning.

as thumb up/down) and *show more/less* (arrow up/down): this design allows users to separate social reactions (*like/dislike*) and algorithmic control (*show more/less*), making users aware of the difference between showing appreciation for a content and choosing to see more of it. This approach, first proposed in a mock-up by the Panoptykon Foundation and People vs Big Tech (2023), aligns with the user empowerment principle of the DSA, as it attributes cognitive responsibility to users.

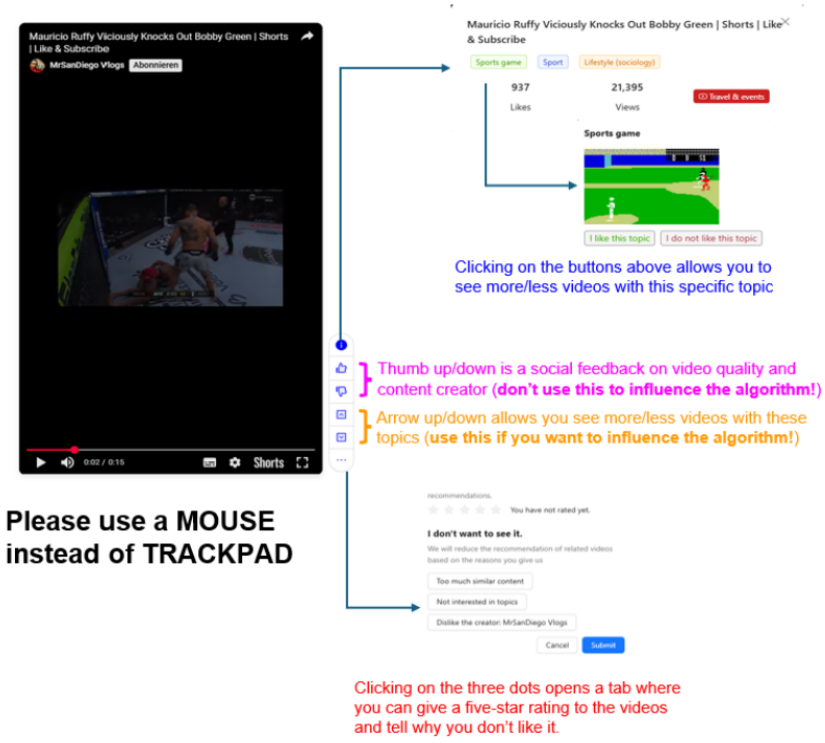


Figure 12: User information page with explanation of the feedback bar on the right side of the video.

The information button at the top of the bar opens the tab with the video title and the three most relevant topics and YouTube category associated with it. By hovering over a topic tag, the user can increase the rating of that topic by clicking *I like this topic*, or decreasing it by clicking *I do not like this topic* (in this case, social feedback and algorithmic control still coincide). These buttons allow users to change the weight of a single topic, while clicking on *show more* or *show less* will increase or decrease the score of each topic associated with the video. Clicking on the three dots at the bottom of the feedback bar opens another tab where the user can find features already present in the previous version of the interface: five-star rating (selecting five stars corresponds to *show more*, while one star to *show less*); explanation menu (whose structure is explained in Table 5; and dislike reason (featuring the options "Too much similar content", "Not interested in topic" and "Dislike the creator").

On the left side of the interface, the control page has to subpages that can be changed through a selector. The "Content Choice" page (figure 13) features controls already available in the old version, i.e. the slider for selecting the proportion between personalized (topic-based) and non-personal (most popular) video recommendations in each round, and the pie chart showing the user's highest-rated topics. The user can change the proportional probability of each topic appearing in the round of recommendations through sliders. In the preliminary preference elicitation, users can select up to five preferred topics or up to three video categories, reflecting the ones embedded in each Short retrieved from YouTube Data API: the selected topics, or the topics included in the selected categories, will be displayed on the pie chart in equal proportion when the user accesses the interface for the first time. The technical updates of the new version concern two main aspects: firstly, the topic rating based on user implicit or explicit feedback is reflected in a real-time change to the topic slider and the corresponding section of the pie chart; secondly, only up to five of the highest-rated topics are explicitly shown, while the rest is grouped in an "Other topics" section of the pie chart that does not have a corresponding slider. These updates aim at increasing the clarity of the pie chart and the accessibility of these controls, reducing the confu-

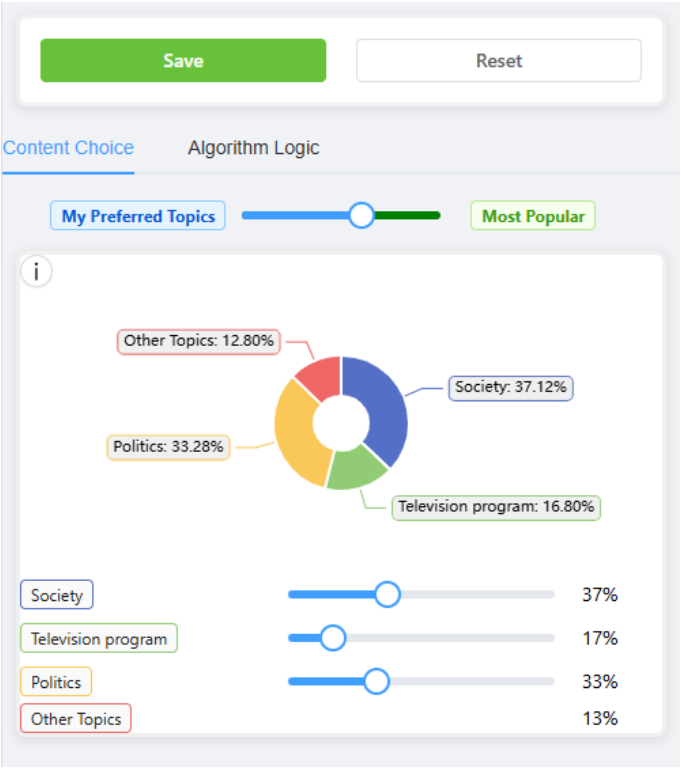


Figure 13: The "Content Choice" page.

sion caused by too many topic names and sliders. The fact that the pie chart and sliders evolve in real time based on user feedback may also contribute to increase users' feeling of empowerment.

The "Algorithm Logic" page (figure14) includes advanced options enabling a granular control of different algorithmic features of the platform. First, while the "Personalization slider" of the previous version keeps the same function, it has been renamed in a more user-friendly way: from "Purely exploitative" to "My Preferred Topics" on the one side, and from "Purely explorative" to "Explore Additional Topics" on the other side. There is, however, an additional slider enabling users to

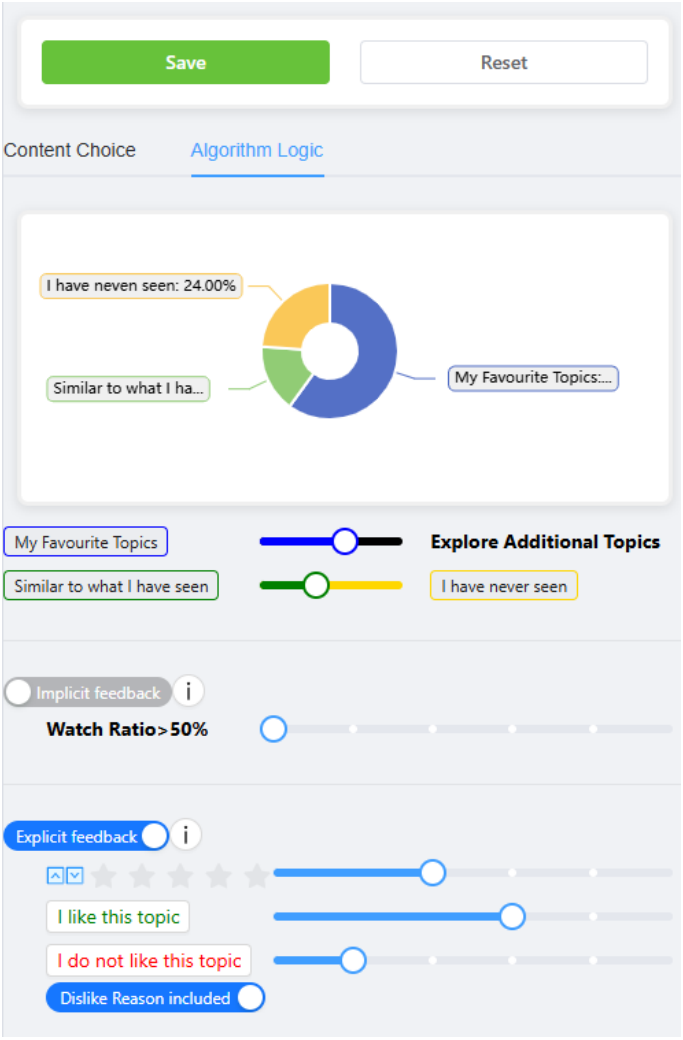


Figure 14: The "Algorithm Logic" page.

determine the composition of the explorative part of their feed: in particular, they can choose proportionally to see videos similar to what they have seen (although not among the top ten highest-rated topics) or that

they have never seen. The proportion of exploitative and explorative recommendations, and the further division within the latter, are showed in a pie chart. In the lower part of the page, there are two buttons allowing users to switch on or off the use of implicit and explicit feedback data in the RS. Implicit feedback consists in watch time, which is retained as an indicator of preference when the user sees at least 50% of a video; explicit feedback includes all the options featured in the feedback bar described above. When the switch is on, advanced users can choose the level of impact that each feedback feature has on the algorithmic rating of topics on which recommendations are based. In this way, users can decide whether and how the data shared during their interaction with the platform will be used for recommendations. This new control option is enabled by the introduction of a weight parameter, W, whose functioning is explained in the next section. Finally, the dislike reason, which works as an independent feedback system and does not directly impact the user profile, is presented as a switch that can be turned on or off if deemed relevant for consideration by the algorithm.

The updated interface refines the implementation of proportionality, granularity and diversity compared to the first prototype. Proportionality and granularity are increased through the new switches and weight sliders that let users decide whether, and to what extent, implicit and explicit feedback types should influence their profile: in other words, users can now modulate how strongly the system reacts to their behaviour, not just how many personalized videos they will see. Moreover, for what concerns granularity, social reactions (like/dislike) are separated from algorithmic control (show more/show less) in the feedback bar. As regards proportionality, the "Content Choice" and "Algorithm Logic" pie charts respectively show, in real time, the proportion of topics featured in the feed and the impact of the algorithmic settings chosen by the user. Finally, diversity controls increased with respect to Chapter 6, as users can also adjust the balance between revisiting less-liked topics and discovering unseen ones, and observe how these settings lead to a broader or narrower spread of topics in the feed. Together with the information buttons, these additions help connect interface actions to so-

cially salient goals - such as improving algorithmic awareness and avoiding over-exposure to certain content - while also making the underlying trade-offs more observable.

7.2.3 Algorithmic Aspects

The platform recommendations rely on the same algorithms described in Chapter 6 [4]. As part of the initial preference elicitation, users can select up to five topics or three video categories they are interested in. In this version of the platform, both topics and categories reflect the ones embedded in YouTube Data API, rather than being modelled ex post as before. A video category is a classification for videos that exists both as an independent object in the YouTube Data API and as an attribute for "search" and "videos" objects. Each video has a unique video category, but there can be overlaps of topics across different video categories. If a user switches to the "Category" tab during preference elicitation and selects their preferences for one or more video categories, all topics within those specific video categories are included in the user initial profile. When multiple video categories are selected, the creation of the user profile for each individual video category is treated similarly. In Equation 7.1 each category is weighted by $\frac{1}{n}$, where n represents the number of video categories calculated, and $p_v(i)$ corresponds to the topic distribution of an individual selected video category:

$$p_0 = \frac{1}{n} * \sum_{i=1}^{n} p_v(i). \tag{7.1}$$

The feedback mechanism is initiated by filtering user interactions based on two primary criteria: the user viewed over 50% of the video's duration (implicit feedback), or the user provided a rating for the video or topic (explicit feedback). Subsequently, topic distributions for each video deemed relevant are derived. The back end operations associated with this functionality entail extensive validation of input from the frontend: the system retrieves the duration of each video to compute the

[4]Only the changes occurred between the original and latest version are reported here: all the other technical specifications are explained in Chapter 6.

watch ratio, which results in redundancy as the frontend similarly verifies whether the user has rated the video.

Following this initial processing, the feedback algorithm takes into account when users click on *I like this topic* or *I do not like this topic* for the topics listed in the detailed pop-up in the video component. Moreover, in the baseline setting, clicking *show more* increases the distribution score of a particular topic by 0.1, while clicking *show less* reduces the likelihood of that topic being shown in the future by 0.5. The subsequent phase involves updating the topic categories, characterized by the computation of the most liked topics, which are identified as the ten topics with the highest user preferences. The topics rated but not among the most liked are classified accordingly, while the remaining topics are categorized as unrated. The scores for these topics are recalculated in accordance with the existing model. Finally, the user profile timestamp is updated to ensure a reference point for future function calls. Additionally, the mechanism for classifying the "dislike reason" is structured as a selection menu. When a user selects a specific dislike reason, the following actions are implemented:

1. When "Too much similar content" is selected, the user's exploration coefficient is decreased by 0.1.

2. When "Dislike the creator" is selected, both the video ID and the channel of the content creator are recorded in the user's profile. Videos from disliked creators will not appear again, but the list of disliked creators is not recoverable by the user.

The new version offers more granular interactions that enable users to gain a deeper understanding of RSs at the algorithmic level. To enhance the level of algorithmic control, a "weight" parameter (W) was added to replace the previously hardcoded constants in the original equation 6.1, allowing users to express their preferences regarding the impact of each interactive functionality on the RS. For each specific feedback functionality, the new parameter replaces the preselected value as follows:

1. ***I like this topic***: This action increases the distribution score of a particular topic by the following amounts (with $+$):

0 (no impact: N) / 0.1 (very low impact: VL) / 0.2 (low impact: L) / 0.3 (medium impact: M) / 0.4 (high impact: H) / 0.5 (very high impact: VH).

2. ***I do not like this topic***: This action reduces the likelihood of that topic being shown in the future by (with $*$):
   1 (N) / 0.8 (VL) / 0.6 (L) / 0.4 (M) / 0.2 (H) / 0 (VH).

3. ***Show more/Show less*** (corresponding to the *like/dislike* buttons of the original version):

   $$f(I, P) = (1 - W) * p_{\text{previous}} + W * \left[-2, -1, 0.5, 1, 2\right]^T * r_{\text{rate}} * v$$

   where $W = 0$ (N) / 0.1 (VL) / 0.2 (L) / 0.3 (M) / 0.4 (H) / 0.5 (VH)

4. **Watch time** (implicit feedback):

   $$f(I, P) = (1 - W) * p_{\text{previous}} + W * r_{\text{watch}} * v$$

   where $W = 0$ (N) / 0.1 (VL) / 0.2 (L) / 0.3 (M) / 0.4 (H) / 0.5 (VH)

In the equations above, p_{previous} denotes the user's current topic preference vector before the new interaction is processed, while $f(I, P)$ represents the updated vector after incorporating that interaction. As a result, new feedback is blended with the previous profile rather than replacing it outright. The rationale is to strike a balance between stability and responsiveness: older interactions continue to shape recommendations, but recent explicit and implicit feedback (including watch time) can gradually shift the profile when the user's interests change. While in the original version user feedback had a constant coefficient of impact on the algorithm, the new version allows the user to select such impact coefficient among pre-defined options. From this perspective, intervening on control options available on the interface can directly influence the underlying RS, as foreseen by recital 70 DSA.

7.3 User Study

The updated version of the platform was tested with 20 users to evaluate whether the improved design and functionalities informed users' un-

derstanding of the recommendation process and perception of empowerment, especially concerning content exposure and diversity and well-being. Given the preliminary nature of this study and the low number of recruited users, statistical significance will not be taken into account; the analysis will rather focus on descriptive statistics on the user sample. Participants were recruited on Prolific [5] based on the following criteria: 1) residing in the European Union and 2) using YouTube. These criteria ensure that participants are both impacted by the DSA and familiar with YouTube Shorts. Participants were then divided into two sets of 10:

- Set A interacted with the platform and filled the final survey autonomously, without receiving guidance apart from the information text at the beginning of the study. Instructions were given in textual form via the study start page (Figure 15) and a pop-up information page with a walk-through of the interface (Figure 16).
- Set B were guided at the beginning and throughout their interaction with the platform by the researcher, who could provide additional clarifications or ask for explanation about user choices. The same textual and graphic instructions were available to them as well (Table 9 describes the workflow of guided participants).

The split between unguided (set A) and guided (set B) participants is meant to capture two realistic modes of using controllability features. Prior work on "algorithmic literacy" suggests that users who receive explicit support in understanding how feeds are curated and how their actions affect recommendations tend to feel more in control and to interpret content exposure more critically (Eslami et al., 2015). At the same time, research on user-controllable personalization shows that the effectiveness of control interfaces depends not only on the available widgets, but also on how their logic is explained and framed to users (Parra and Brusilovsky, 2015; Millecamp et al., 2018). Building on these insights, set A operationalizes a realistic platform use scenario, where users explore controls largely on their own, while set B approximates a media-literacy

[5]Participants gave their informed consent to participate in the study.

1. Use your Prolific ID as your username to register on the platform and fill the initial survey.
2. Select EITHER topics OR categories to inform the system about your initial preferences: these will be used for the first recommendations you receive.
3. The user study will consist of 7 rounds, each showcasing 10 short videos. You should provide feedback on the videos using the buttons on the right of each video**, and interact with relevant control features on the "Content Choice" and "Algorithm Logic" page displayed on the left side of the interface**.
4. At the end of the 7 rounds, you will be asked to complete a survey to assess your user experience on the platform. Please make sure to complete all the survey with attention to detail.

Notice:

1. Please do not refresh your browser or close the session and return during the study. Otherwise, the rounds will be counted again.
2. For the best experience, use a mouse instead of a trackpad during the study.
3. The most suitable screen size for the study is 24 inches. However, it can also be used on 14, 16, 27, or 32-inch screens. Please adjust the zoom level to ensure the display is properly sized.
4. Note that the scroll wheel on your mouse is disabled for other functionalities other than switching between videos on the short video platform. If you need to zoom in or out during the study, please use your browser settings.

Please register or log in first to start the survey

Figure 15: Study start page with the instructions shown to participants before the study.

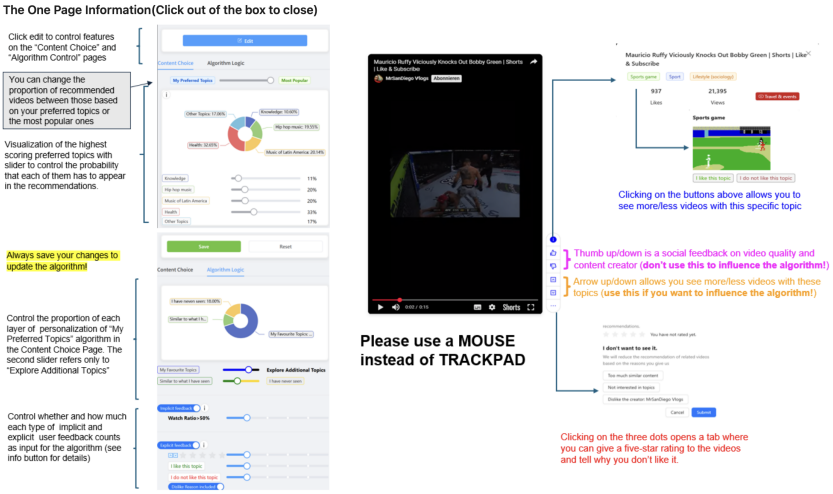


Figure 16: Pop-up information page with a walk-through of the interface explaining each feature and control option.

Time	Task	Description
2 min	Introduction	Participants are introduced to the study workflow, following the guidelines provided on the front page.
5 min	Registration	Participants complete the pre-experiment survey and preference elicitation phase.
10 min	Interface Overview	The interface features and their relationship with the RS are explained in detail.
30 min	User Interaction	Participants are invited to watch the videos and interact with the interface features. In case of doubts, they can ask questions to the researcher, who may intervene to provide clarification or check user understanding.
5 min	Survey	Participants are asked to complete the final survey and may be asked clarificatory questions on their answers.
5 min	Follow-up Questions	Participants may receive clarifying questions related to the their choices in interacting with the platform and responding to the survey.
3 min	Feedback	Participants are invited to share any suggestions or complaints about the system or the study process.

Table 9: Timeline and procedure of the guided user study

or assisted-use scenario in which guidance is available throughout the interaction.

Participants' task unfolds as follows. After registration, they encounter the same pre-experiment survey about social media use appearing in the old version. Then, a pop-up for preference elicitation prompts them to select up to five topics or three categories they are interested in and that will be featured in the "Content Choice" page. As soon as they access the platform interface, a full-screen information slide on how to use the various control options opens (and remains accessible throughout all the study by clicking on the "i" icon in the top left corner of the interface).

The control page on the left side is left open by default, but can be collapsed by the user. After clicking the start button of the first video, participants are exposed to seven rounds of ten recommendations each, for a total of 70 Shorts. Participants are free to choose if and when to use control tools, scroll to the next video or pause the video stream. If the participant does not scroll or pause the video, the feed will automatically proceed to the next video, as happens in the real-world platform. After the last recommendation, the final survey pops up: once the user fills it, the task is completed. The estimated time needed for completing the study is around 50 minutes: set A users took between 20 and 81 minutes, while set B took between 50 and 65 minutes.

The final survey has been redesigned to focus on the perceptions of empowerment and wellbeing, which are the most impactful from the perspective of the DSA. In particular, the first 11 questions refer to the platform in general, while the subsequent 17 (some of which include subquestions, for a total of 29) concern each specific interface feature. Both general and feature-specific questions examine the same dimensions of the user experience: algorithmic awareness, control over content exposure and recommendation process, diversity, protection of minors[6] and vulnerable people. These dimensions expand and specify those related to perceived controllability, perceived transparency and digital wellbeing in the user study of Chapter 6. For the complete list of survey questions, see Table 10.

Q#	General questions
1	I understand how the recommendation process works, i.e. how it impacts the order and topics of the videos displayed on the interface.
2	Using this interface made me feel empowered to directly influence the recommendation process.
3	I am satisfied with my ability to actively control the video recommendations.

[6] All participants were older than 18. Survey questions related to the protection of minors were aimed at connecting participants' perception of the proposed interface design with the rationale of the DSA guidelines on protection of minors (Commission, 2025).

Q#	(continued)
4	After interacting with this interface, my ability to recognize how recommendations direct my attention to certain categories of content has improved.
5	I think that using the control features provided can make user experience more diverse and meaningful.
6	I am satisfied with my overall experience with the platform.
7	If three plus five is eight then click disagree.
8	My expectations on how the recommended videos would change after using the control features corresponded to what I actually saw on the interface.
9	The use of control features increased my ability to avoid content that I prefer not to see.
10	The use of control features increased my exposure to the content I prefer to see.
11	I think the type of control features provided can help protect minors and vulnerable people online.
	Feature-specific questions
1	Separating *Like/Dislike* from *Show more/Show less* makes me feel more in control of the recommendations.
2	*Show more* increases my exposure to content I prefer.
3	*Show less* enhances my ability to avoid content I do not want to see.
4	I think *Show more/Show less* can make user experience more diverse and meaningful.
5	*I like this topic* increases my exposure to content I prefer.
6	*I do not like this topic* enhances my ability to avoid content I do not want to see.
7	I would prefer using *Five-star rating* instead of clicking *Show more/Show less*.
8	Choosing the "Dislike reason" enhances my ability to avoid content I do not want to see.
9	The slider *My Preferred Topics – Most Popular* on the "Content Choice Page" made me feel empowered to directly influence recommendations.
10	Seeing how the pie chart on the "Content Choice" page changes based on my actions motivates me to engage with various feedback features (e.g., *Show more/Show less*, *I like this topic* or *I do not like this topic*).

Q#	(continued)
11	The "Content Choice" page: a) helps me understand how the recommendation process works and how it impacts the order and topics of the videos displayed on the interface. b) can make the user experience more diverse and meaningful. c) should be implemented on mainstream short video platforms.
12	The "Algorithm Logic" sliders: a) allow me to diversify the topics of the videos I see. b) allow me to choose whether the recommendations will focus on my favourite topics or on exploring additional topics, even some that I have never seen before.
13	The recommended videos after adjusting the *Implicit feedback* and *Explicit feedback* switches align with my expectations regarding how the suggested content would change.
14	The "Algorithm Logic" page: a) made me feel empowered to directly influence the recommendation process. b) should be implemented on mainstream short video platforms.
15	If nine minus three is six, please click neutral.
16	Controlling the percentage of *My Preferred Topics* as they appear in the pie chart: a) improved my understanding of why I receive certain recommendations b) made me feel empowered to directly influence the recommendation process. c) increased my ability to avoid content that I do not want to see d) increased my exposure to the content I prefer to see e) should be enabled also on mainstream short video platforms.
17	Controlling the impact of each feature under *Implicit feedback* and *Explicit feedback* on the "Algorithm Logic" Page: a) motivated me to interact with different feedback features. b) made me feel empowered to directly influence the recommendation process. c) increased my ability to avoid content that I do not want to see. d) increased my exposure to the content I prefer to see. e) should be enabled also on mainstream short video platforms.

Table 10: Survey questions on platform control features.

7.3.1 Results

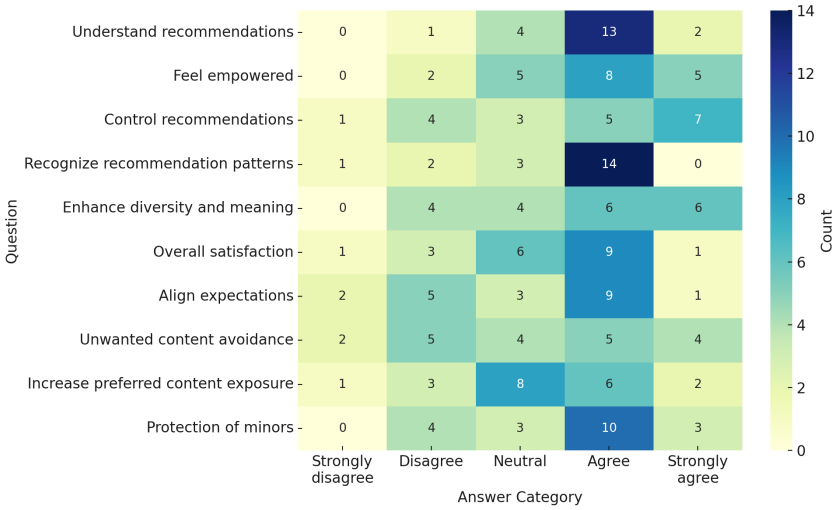


Figure 17: Heatmap showing the frequency of answer category for the general survey questions.

Having conducted the study in both an unguided and guided mode, it is possible to account for both the aggregate results and the results for sets A and B separately. The heatmap in Figure 17 shows the aggregate results for the general questions[7]. As can be seen, the cells with highest frequency refer to two questions about algorithmic awareness: indeed, 70% of users responded ”agree” to Q1 and 65% to Q4. Overall, the column with highest frequency is ”agree”, which in two cases is surpassed by ”strongly agree” (Q3) and ”neutral” (Q10), and in two cases matched by ”strongly agree” (Q5) and ”disagree” (Q9). The cases in which user

[7]The order of questions in this and the next heatmaps follows that of Table 10. The two attention checks, one per each section of Table 10 (question 7 and 15, respectively) are not showed in the heatmaps, as all participants passed the attention checks and including the corresponding rows would reduce the usefulness of the frequency visualization. The question text is summarized through a label in every heatmap. From now on, to refer to a specific question, the capitalized initial ”Q” followed by the relative number will be used: numbers always refer to those displayed in Table 10.

preferences are less concentrated concern the satisfaction with one's ability to control recommendations (Q3) and the perception that the use of control features can make user experience more diverse and meaningful (Q5), steer recommendations away from unwanted content (Q9) or increase exposure to preferred content (Q10). For the first five questions and the last one the majority of answers are "agree" and "strongly agree", indicating a positive perception of the impact of the platform design on awareness and understanding, user control, UX diversity and protection of minors/vulnerable subjects. For two questions, concerning the overall satisfaction with the platform (Q6) and the fulfilment of expectations regarding how recommendations would change after using control options (Q8), the amount of positive responses equalled that of neutral and negative ones. Only Q9 and Q10 had more neutral and negative responses (55% and 60% respectively) than positive ones, indicating that just more than half users did not perceive an impact of the use of controls in decreasing exposure to unwanted content and vice versa.

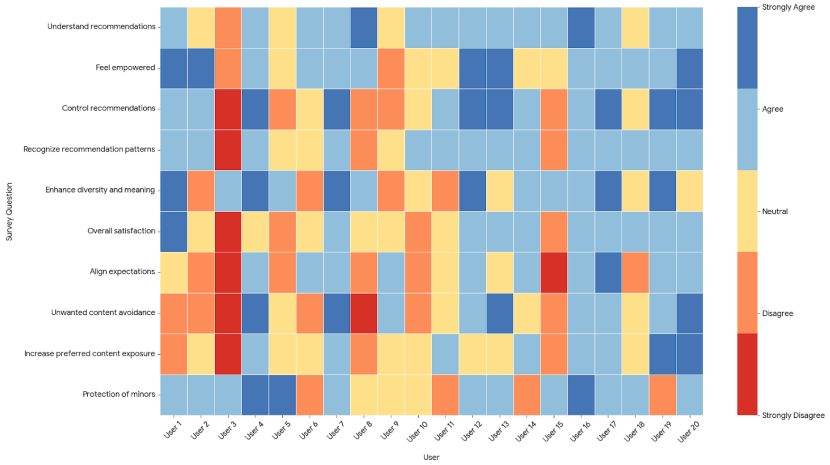


Figure 18: Heatmap showing the distribution of answers to the general survey questions across respondents. Users 1-10 belong to set A, while the rest belong to set B.

The heatmap in figure 18 shows the distribution of answers across

participants: users 1-10 belong to set A, while users 11-20 belong to set B. As can be observed, most negative (”disagree” and ”strongly disagree”) and neutral answers come from set A, with only three out of ten users giving a majority of positive answers (”agree” and ”strongly agree”). Looking at set B, the opposite situation emerges: seven out of ten users gave a majority of positive answers. Moreover, among the non-positive responses given by each set, the following pattern emerges: for set A the number of negative answers exceeds that of neutral ones, while for set B the opposite is true. These results indicate a diverging pattern of responses between the two sets: respondents from set B responded more positively than those from set A to all questions. The questions with a majority of positive answers in each set of users (considered separately) are Q1, Q2, Q5 and Q11: overall, the system is perceived as enabling users' understanding (75% positive answers)[8] and feeling of empowerment (65% positive answers) while potentially enhancing diversity (60% positive answers) and protection of minors (65% positive answers). 70% of users reported an improved ability of how recommendations influence their attention after interacting with the interface (Q4). Considering negative answers, the expectations on how recommendations would change after using control options did not match reality for 35% of users, while the same proportion felt that the use of control features did not increase their ability to avoid unwanted content.

These results are further specified by the responses to feature-specific questions (figure 19). In this case, the most frequent answer to the majority of questions (22 out of 28) is ”agree”. Moreover, positive answers are more than 70% for half of the questions. In particular, the new functionality *show more/less*, separated from *like/dislike*, made 80% of users feel more in control of recommendations, also for what concerns increased exposure to preferred content and ability to avoid unwanted content. The real-time changes in the pie chart on the ”Content Choice” page motivate 75% of users to engage with various feedback features. Moreover, controlling the percentage of preferred topics in the pie chart make three quarters of respondents understand why they receive certain recommen-

[8]Percentages of positive and negative answers are calculated over total respondents.

dations, while it enhances the feeling of empowerment and increases exposure to preferred content for 70% of them. 80% of users think that the "Content Choice" page helps them understand how the recommendation process works, while 70% believe it can make user experience more diverse and meaningful. Similar results characterize the questions about the "Algorithm Logic" page, by which 65% of users feel empowered. Specifically, most respondents think that the sliders on this page allow them to diversify the topics of videos they see (80%) and to choose the proportion of exploitative and explorative recommendations effectively (75%). 75% of users feel empowered by the possibility to control of the impact of each feedback feature on recommendations, while 70% perceive that it increases exposure to preferred content.

Interestingly, although many features of the "Content Choice" and "Algorithm Logic" pages were appreciated by most users, the four questions on whether these pages and some of their options (i.e., controlling the percentage of topics and the impact of feedback type) should be implemented also on mainstream short video platforms did not yield a majority of positive answers. In particular, the most frequent answer to these question is "neutral" which, in the case of Q16(e) (control of topic percentage") is equalled by "strongly agree". Half of the users answered positively to Q11(c) ("Content Choice" page), while for other three questions non-positive answers are the majority. Another case in which "neutral" responses are the majority (60%) concerns Q13 about the expected impact of turning on or off the *implicit/explicit feedback* switches: this highlights a discrepancy with the mostly positive answers to four sub-questions 17. The only case in which the most frequent answer is negative ("disagree") is Q7, concerning the preference of five-star rating over *show more/less*. Considering the concentration of positive answers to Q1-4, this indicates a common preference for *show more/less* as a direct way to control recommendations.

The heatmap in figure 20 shows the distribution of answers to each feature-specific question across respondents. As can be observed, answers are less polarized than those to general questions displayed in figure 18. While positive answers are still more frequent for set B, in which

just one user gave mostly non-positive responses, six out of ten users in set A gave a majority of positive answers. Comparing the answering pattern of the same user across the two heatmaps underlines their degree of consistency. For example, user 4 shows a consistently positive answering pattern to both general and feature-specific questions; similarly, user 15 responded negatively to the most of the general and feature-specific questions. Instead, user 3 answered "strongly disagree" to most general questions and "neutral" to the majority of feature-specific questions: insights on discrepancies like this may emerge through the *ad hoc* analysis of user interaction logs performed in the next section.

7.4 Discussion

The differences in response patterns between set A and set B, especially evident in the first part of the survey (Figure 18), may be attributed to the researcher's guidance. In fact, positive answers from set B (the guided users) are more than positive answers from set A (the non-guided users), especially for general questions. While the task was the same for sets A and B, the constant availability of guidance, also in the form of proactive prompts aimed at stimulating self-reflection and understanding, may have motivated a higher appreciation for control features by the latter. Apart from stimulating understanding, the researcher's presence might have caused social desirability bias in set B respondents, who might have conformed their feedback to what they assumed would be the researcher's expectation. However, the evidence gathered through unstructured questions in the case of set B, as well as the analysis of the interaction logs for set A, draws a more complex scenario than this dichotomic one.

In fact, the comments made by set B users shed light on the patterns emerging from the results. First, none experienced the granularity of controls characterizing this experimental platform before. Most respondents argued that, while they might use control options similar to those experienced in this context if available on mainstream platforms, the majority of "standard" social media users would not use them due to the

level of cognitive effort required for an effective controllability. User 15, who gave mostly negative responses to both general and survey-specific questions while engaging with almost every control feature, was not satisfied with the outcome ("strongly disagree" about the correspondence between expectations and reality) despite considering controllability important and in fact arguing in favour of the usefulness of the controls for the protection of minors. Users 14 and 19, who instead answered positively in almost every instance, argued the opposite on that matter, as they claimed that controls are not user-friendly enough for minors or could even be used to access harmful content more directly through voluntarily sought feedback loops and heavily exploitative recommendations. The assumption that most users would not use granular control features translates to a lack of incentives for large providers to change the design of their RSs in the direction of controllability. User 13, who engaged extensively with control features of different complexity, reflected on the implications of the attention economy driven by platforms and argued that algorithmic education is a prerequisite for informed and effective user control.

For what concerns set A participants, insights can be drawn from their recorded interactions, whose average amount is 84 per user. The previously identified mismatch for user 3, who answered "strongly disagree" to six out of ten general questions and "neutral" to most feature-specific questions, can be explained by their limited use of controls: out of 17 total interactions, just the final two concerned the high-level control pages, while all the rest were on the feedback bar. Therefore, this user might have experience a high dissatisfaction with the system, but, not having tried most features, they answered "neutral" to the majority of feature-specific questions. User 4, who answered positively to almost all questions, interacted mostly through the feedback buttons on the right of the video, sometimes modified the topic sliders under the pie chart but never used more granular controls, such as those on the "Algorithm Logic" page. After expressing negative feedback multiple times throughout 60 interactions, they were eventually exposed to more interesting content, as signalled by *like* and *show more* towards the end of

their logged interactions, which might have turned them towards positive answers in the survey (although the overall experience is "neutral").

User 1, who took 81 minutes to complete the study for a total of 275 interactions (the highest recorded completion time and interaction number, respectively), tried every type of control options and gave nuanced answers. While disagreeing that the control features help increase exposure to preferred content and avoid unwanted content, they answered positively to all the general questions concerning understanding (Q2), awareness (Q4), feeling of empowerment and control (Q2, Q3), diversity/meaningfulness (Q5), overall satisfaction (Q6) and protection of minors (Q11). User 8, the only other participant whose logs show more than 200 interactions, expressed negative perceptions with regard to content exposure, expectation alignment, satisfaction with control and awareness on how recommendations direct their attention; positive answers concerned understanding, feeling of empowerment and diversity. Despite a similar level of engagement with the system, user 1 and 8 reported different degrees of control and awareness: while both of them understood how the recommendation process works and felt empowered by the interface, their reported satisfaction with controls and grasp of how recommendations steer their attention diverge.

The results discussed above, clarified by users' qualitative insights, allow to outline some findings for Chapter 7[9]:

- **First, guided users felt more empowered, in control and aware of the influence of recommendations than non-guided ones.** While this may be partially explained through social desirability bias, the researcher's role of guide/educator might have resulted in an increased algorithmic literacy. At the same time, negative answers by non-guided users cannot be solely attributed to superficial engagement with the platform, given the high number of interactions by some respondents in this set. As highlighted in Section 7.3, the

[9]Given the small user sample, this study does not aim to provide statistically significant observations. It should rather be considered as an exploratory inquiry on the potential impact of the DSA provisions on RSs on the design and controllability of short video platforms.

researcher's guidance can also be interpreted as an effort to educate users (Eslami et al., 2015). In this sense, the more positive responses from set B illustrate what could be achieved in an algorithmic literacy setting where sustained guidance is available; however, such intensive, one-to-one support is unlikely to be feasible in a social media service and should therefore be understood as an upper bound on what interface-embedded explanations and onboarding flows might accomplish for the transparency and user empowerment goals of the DSA.

- **Second, across both sets, most users reported understanding how the recommendation process works, feeling empowered to directly influence the RS, having a more diverse and meaningful experience through control features**, which, they believe, can help protect minors and vulnerable people online. While most users appreciated controllability as a way to improve their experience in online platforms, the effectiveness of controls for specific aims is debated: in fact, just over half of the participants did not experience a sensible reduction of unwanted content or a neat increase of preferred content after using the control features. Therefore, the combination of feedback-based and algorithmic controls improved perceived understanding and empowerment, but it did not fully deliver on users' expectation of being exposed to content more aligned with their preferences.

- **Third, updates to the platform resulted in higher user appreciation for both low- and high-level control options, while enhancing proportionality and granularity.** New features like the *show more/less* buttons and the *implicit/explicit feedback* impact sliders were positively received by more than half of users. Cognitive load was not reported as a hindrance by individual participants, but some expected that the average user would feel otherwise.

7.5 Limitations and Conclusion

While the updated design and functionalities of the interface allowed to overcome some shortcomings described in Chapter 6, this second user study is still exploratory in scope and has tangible limitations. The researcher's guidance, marking the difference between the two sets of participants, can framed as a proxy for interventions fostering algorithmic literacy, but it does not represent a realistic baseline for common use of short-video platforms; in particular, the study does not disentangle which aspects of the guidance (e.g. walk-through of control options, reminders about their effects) are most effective or scalable beyond a prototype setting. Due to the limited amount of qualitative insights, nuanced participants' perceptions were not observable in most cases, and, when available, might have been affected by social desirability bias. The study has not been conceived to provide definitive evidence about the effectiveness of such measures in real-world platforms.

In conclusion, while this prototype cannot be compared to real-world applications in terms of satisfaction of user preferences and effectiveness of recommendations, its intent of fostering user awareness and empowerment was generally acknowledged by participants. The updated interface features represent a further application of the principles of proportionality, granularity and diversity of Chapter 4 and, as such, may constitute a design proposal for the standardization of the DSA requirements on the transparency and controllability of RSs. As a promising development in this direction, in 2025 TikTok implemented a feature called "Manage Topics" through which "users can customise how often content related to over 10 topics is recommended in their FYF" (For You Feed), as the provider details in its 2025 DSA systemic risk assessment report TikTok, 2025: in particular, "users can move a slider to adjust how much they want to see of each topic". This type of topic-based control is very close to the one presented in Chapter 6 (Figure 4), developed before TikTok's implementation. Such advancement demonstrates that user empowerment through explicit topic-based control of RSs is already part of platforms' compliance with the systemic risks obligations of the DSA.

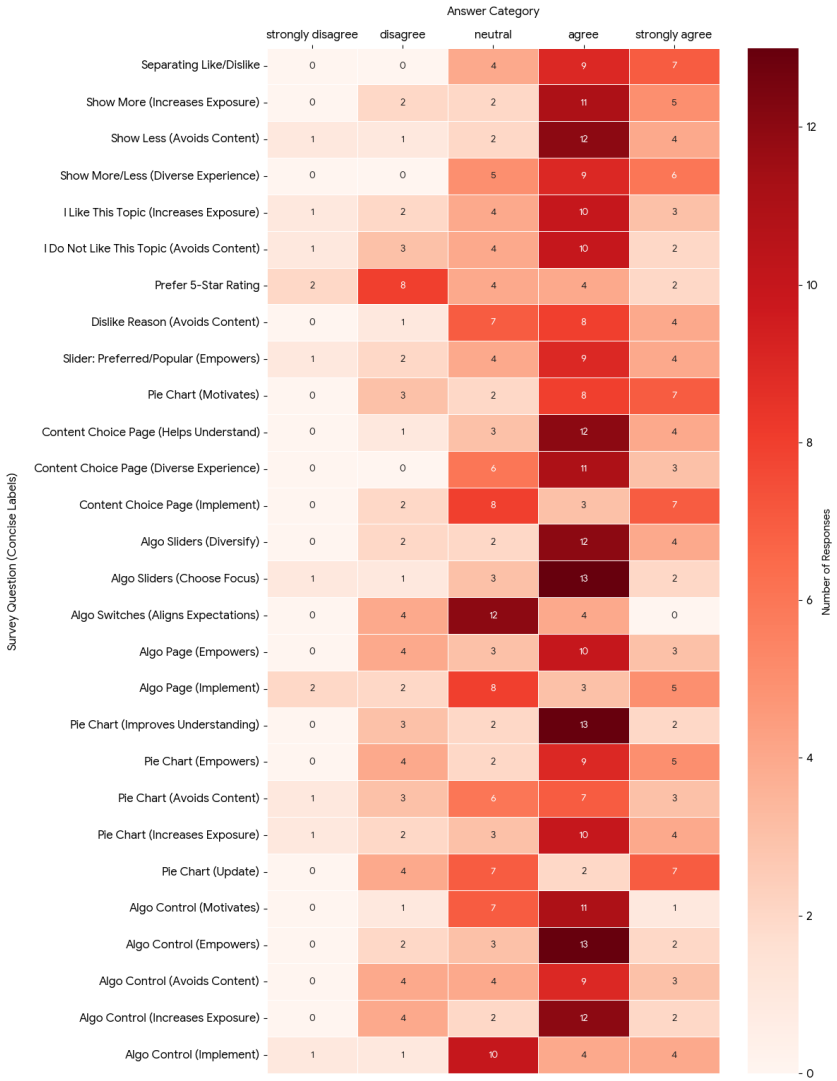


Figure 19: Heatmap showing the frequency of answer category for the feature-specific survey questions across all participants. The first eight questions from the top concern feedback-based controls, while the rest concern the more granular algorithmic controls in the Content Choice and Algorithm Logic pages.

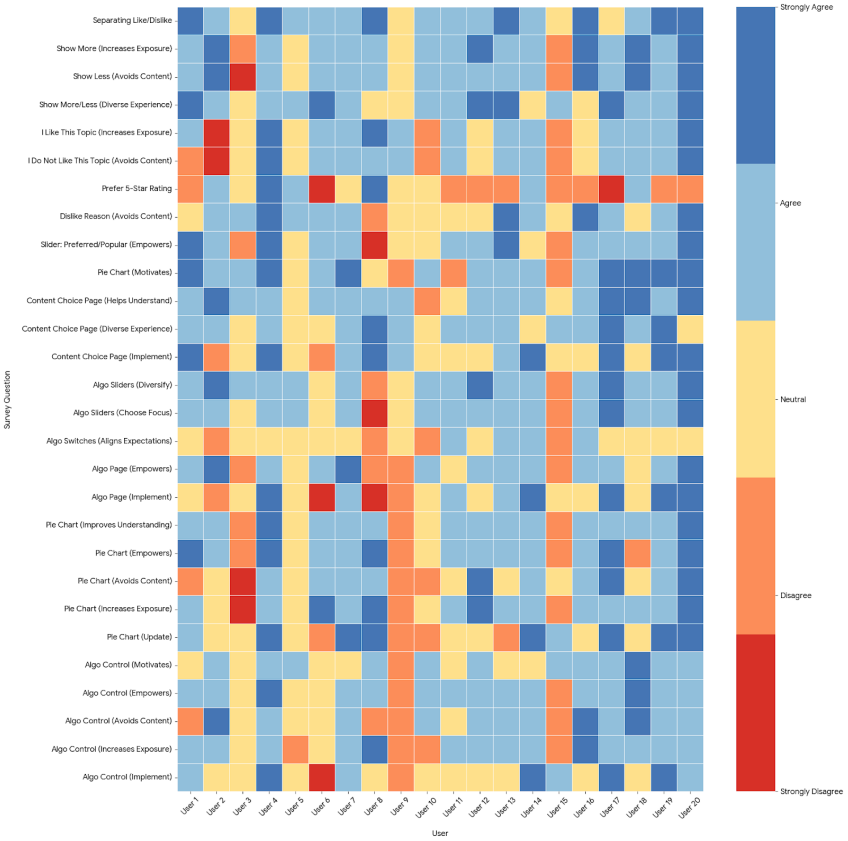


Figure 20: Heatmap showing the distribution of answers to the feature-specific survey questions across individual respondents. Each column corresponds to a participant (users 1-10 belong to set A, while the rest belong to set B), each row to a question. The first eight questions from the top concern feedback-based controls, while the rest concern the more granular algorithmic controls in the Content Choice and Algorithm Logic pages.

Chapter 8

Conclusion

This chapter concludes the thesis by outlining how the insights provided by each pillar - philosophical, legal and empirical - integrate a perspective on how the influence of platforms on users can be redirected towards the objectives of the latter through the transparency and controllability of recommender systems envisioned by the Digital Services Act. It also addresses the limitations of the interdisciplinary approach of the thesis.

8.1 Main Contributions

For an effective application of the user empowerment principle emerging from the DSA, the design choices supporting the transparency and controllability of RSs cannot be left to providers, but have to be directed through regulatory guidelines and standard-setting processes. Most of the systemic risks that the DSA seeks to tackle are enabled by RSs, from young users' mental and physical health (Costello et al., 2023) to misinformation (Epstein, Pennycook, and Rand, 2020). However, the legal provisions about RSs remain vague for what concerns the design of user-facing transparency and control features. The scope of their implementation is shaped by inconsistent interpretations, which often tend towards minimal compliance. This dynamic is reflected in DSA-mandated platform audits (Fabbri and Boratto, 2025), which evaluate similar compli-

ance practices differently while disregarding substantial shortcomings in transparency reporting (Figure 21). This has also been recognised by the regulator, who admitted that "the signal to noise in the audit reports is not what we wanted to be" (Agarwal, 2026). As a result, the current state of transparency and controllability in major platforms falls short of ensuring that users "can influence how information is presented to them" through recommendations (recital 70 DSA).

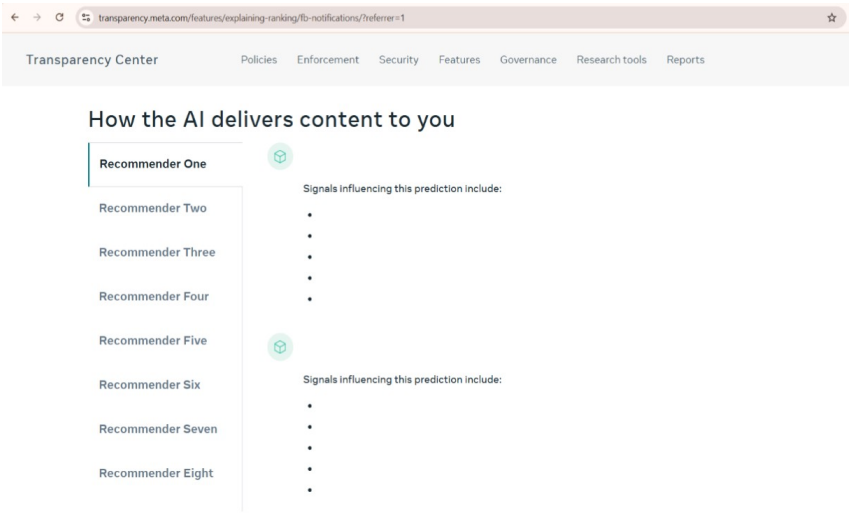


Figure 21: Facebook's system card with the main parameters for its notification RSs appeared blank before the latest update on 24th March 2025.

It is unlikely that platforms will voluntarily redesign interfaces or underlying algorithmic infrastructures that underpin their profitable business models unless compelled by credible enforcement pressure. However, while the enforcement of other EU digital laws - such as the Digital Markets Act (DMA) - has already resulted in various fines for gatekeepers (Moens, 2025), only the first proceeding initiated under the DSA has reached an end in December 2025, with a 120 million euros sanction for the provider of X (Fabbri, 2026). As argued throughout this thesis, two factors help explain this pace: first, the lack of a robust scientific and tech-

nical knowledge base; second, the current geopolitical scenario. With respect to the first, the Commission's need to rely on external expertise to enforce the DSA has been demonstrated through various call for tenders issued by CNECT.F (known as the DSA enforcement team): the tendered tasks span from producing literature reviews (European Commission, 2024a) and technical tools to analyse systemic risks (European Commission, 2024b) to developing "comprehensive user models that capture the complex interactions between users and recommender systems on online platforms" (European Commission, 2025a). The data access for vetted researchers based on art. 40 DSA is another instrument through which the Commission aims to collect scientific evidence to support enforcement. Regarding the second factor, the "ongoing pressure from the Trump administration to challenge the implementation of European digital regulations" (Jahangir, 2025) may have slowed down the pace of DSA enforcement amid EU concerns about transatlantic trade relations (Moens, 2025). In this geopolitical environment, RSs represent a strategic asset for influencing citizens, as underlined by the US government's move to transfer the ownership of TiKTok's US operations to US investors amid concerns that "the algorithm that fuels what users see on the app is vulnerable to manipulation by Chinese authorities" (Madhani, 2025). Nonetheless, neither the lack of a sound scientific base nor ongoing EU-US geopolitical tensions can be directly linked to specific Commission decisions under the DSA, given that the vast majority of documents on the enforcement process are kept confidential (Fabbri, 2025b).

Against such backdrop, my thesis aims to challenge this normative, political and procedural opacity by accounting for the philosophical context, legal application and empirical meaning of user empowerment through transparent and controllable RSs in social media platforms. The three pillars of the thesis - each of which represents a different while complementary lens to observe the problem at hand - collectively offer an interdisciplinary perspective that illuminates the conceptual roots, regulatory formulation and socio-technical application of the user empowerment principle in EU digital law. Specifically, the research questions pertaining to each pillar have been addressed as follow:

- **Philosophical pillar**: Since users can decide whether to consume recommended content but not what content is recommended, RSs derive their influence from users' reliance on them for accessing information. Empowering users therefore means enabling them to exercise their positive liberty by controlling *for what* and *to what extent* they want to rely on the system, as argued in Chapter 2.

- **Legal pillar**: Chapter 3 underlines that, in order to operationalize user empowerment, the DSA provisions on the transparency and controllability of RSs should be complemented by guidelines or standards. The current lack of technical definitions and thresholds leads to inconsistent compliance practices by platforms and opaque enforcement decisions by regulators, as outlined by Chapters 4 and 5. To overcome this lack of clarity and initiate a stand-setting process, Chapter 4 proposes a speculative design framework pairing granular algorithmic choice with diversity- and authoritativeness-aware curation, which could lead platforms to substantiate meaningful personalization and user empowerment.

- **Empirical pillar**: When provided with proportional and granular controls for RSs on the short-video platform described in Chapter 6, most selected users reported understanding how the recommendation process works and feeling empowered to directly influence it while having a more diverse experience. Algorithmic literacy can transform user control from a cognitive burden to an opportunity for empowerment, as argued in Chapter 7.

To address the research question guiding this thesis: **the DSA provisions on the transparency and controllability of RSs can meaningfully empower users only if platforms are required — through regulatory guidelines or standards — to move from an interaction model centred on passive engagement to one that enables users to consciously co-curate their feeds and participate in the governance of RS design**. Under such conditions, a robust implementation of the user empowerment principle would lead to a shift in platform business models: from treating users as passive targets of engagement optimization to recognizing

them as architects of the information flows shaping their online interactions. Value creation would no longer derive from maximising the profit deriving from users' fragmented attention, but from enabling them to articulate their interests and streamline their preferences through a conscious, self-determined interactive experience.

8.2 Limitations and Future Research

The interdisciplinary approach adopted in this thesis has the advantage of addressing its topic from different angles: the first pillar brought about ontological, ethical and political-philosophical contributions; the second pillar included perspectives from EU law and governance, socio-legal studies, philosophy of technology and speculative design; the third pillar featured human-computer interaction approaches. Each disciplinary contribution had an ancillary function: supporting the development of the main argumentative line of the thesis. Every pillar is in dialogue with the others, cross-fertilizing the respective results. This approach is reflected in the content and sources of the chapters, which are mostly based on previously published papers of mine, as declared in Chapter 1. Throughout my doctoral research, I prioritized the publication of articles that could directly feed the development of the thesis. The topics covered by the outlets in which my articles are published reflect their interdisciplinary character: from AI ethics to media studies, from philosophy of technology to computer science.

This choice underlines the aim of my thesis, i.e. to address the meaning, application and implications of the EU legal provisions on the transparency and controllability of RSs. For these reasons, I decided not to include systematic literature reviews, wide-ranging doctrinal or documental analyses, large-scale user experiments: despite their usefulness in clarifying concepts or phenomena related to my research, focusing solely on either of these approaches would have led to account for one dimension, rather than the multi-dimensional spectrum, of the research question at hand. From the selection of legal sources, policy cases and documental evidence, to the ideation and use of a short-video platform

for user studies, I believe that the varied methodological approach of this thesis is consistent with the first stages of enforcement of a digital regulation like the DSA which, due to its socio-technical nature, requires different types of expertise to be applied at the same time.

Further developments of this research should aim at systematizing the findings and approaches used in each pillar: from a philosophical standpoint, examining how the conceptual roots of user empowerment unfold in social theory and, relatedly, how the non-legal terminology related to influence found in EU digital regulations can be meaningful for law enforcement; from a legal standpoint, analysing, through the early evidence emerging from court judgements, auditing and enforcement actions, how platforms change the design of their algorithms and interfaces to comply with regulatory obligations; from an empirical perspective, making use of the DSA data access mechanism for vetted researchers to generate socio-technical insights into how the influence of platforms can be redirected through user empowerment. The integration and systematization of these research streams could result in a new body of knowledge that can be used not only to understand digital phenomena but also to govern them.

Bibliography

Abu-El-Haija, Sami et al. (2016). "YouTube-8M: A Large-Scale Video Classification Benchmark". In: *CoRR* abs/1609.08675. arXiv: 1609.08675. URL: http://arxiv.org/abs/1609.08675.

Adomavicius, Gediminas and Alexander Tuzhilin (2005). "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions". In: *IEEE transactions on knowledge and data engineering* 17.6, pp. 734–749.

Agarwal, Prabhat (2026). *Keynote*. DSA and Platform Regulation Conference. URL: https://hva-uva.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=09c7a3b7-cf0b-47cb-b313-b3d3009d33c4.

Aljukhadar, Muhammad, Sylvain Senecal, and Charles-Etienne Daoust (2012). "Using recommendation agents to cope with information overload". In: *International Journal of Electronic Commerce* 17.2, pp. 41–70.

Amoo, Taiwo and Hershey Friedman (Jan. 2001). "Do Numeric Values Influence Subjects' Responses to Rating Scales?" In: *Journal of International Marketing and Marketing Research* 26, pp. 41–46.

Amsterdam Court of Appeal (2026). *Judgment of the multi-member civil chamber of 10 March 2026 in the case of: META PLATFORMS IRELAND LTD. versus STICHTING BITS OF FREEDOM*. URL: https://www.bitsoffreedom.nl/wp-content/uploads/2026/03/Court_of_Appeal_Amsterdam_%E2%80%93_Judgment__March_10_2026__%E2%80%93_Meta_Ireland_vs__Bits_of_Freedom.pdf.

Ardagelou, Panagiotis and A. Arampatzis (2017). "A Half-Life Decaying Model for Recommender Systems with Matrix Factorization". In: *TDDL/MDQual/Futurity@TPDL*.

Auger, James (2013). "Speculative design: crafting the speculation". In: *Digital Creativity* 24.1, pp. 11–35.

Baier, Annette (1986). "Trust and antitrust". In: *Ethics* 96.2, pp. 231–260.
Bakalov, Fedor et al. (2013). "An approach to controlling user models and personalization effects in recommender systems". In: *Proceedings of the 2013 International Conference on Intelligent User Interfaces*. IUI '13. Santa Monica, California, USA: Association for Computing Machinery, pp. 49–56. ISBN: 9781450319652.
Barbieri, Lorenzo (2024). *What does the X exodus to Bluesky mean for journalism?* The Conversation. Accessed: 2025-11-03. URL: https://theconversation.com/what-does-the-x-exodus-to-bluesky-mean-for-journalism-246296.
Bartmann, Marius (2023). "Reasoning with Recommender Systems? Practical Reasoning, Digital Nudging, and Autonomy". In: *Recommender Systems: Legal and Ethical Issues*. Ed. by Sergio Genovesi, Katharina Kaesling, and Scott Robbins. Cham: Springer International Publishing, pp. 129–145. ISBN: 978-3-031-34804-4.
Baumer, Eric PS, Mark Blythe, and Theresa Jean Tanenbaum (2020). "Evaluating design fiction: The right tool for the job". In: *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pp. 1901–1913.
Beatriz Botero Arcila (2024). *An early win for the transparency measures of the DSA. A comment on Amazon Services v. European Commission (C-638/23)*. Accessed on 16 December 2024. URL: https://dsa-observatory.eu/2024/05/02/an-early-win-for-the-transparency-measures-of-the-dsa-a-comment-on-amazon-services-v-european-commission-c-638-23/.
Berlin, Isaiah (1969). "Two Concepts of Liberty". In: *Four Essays on Liberty*. Oxford: Oxford University Press.
Bits of Freedom (2025). *Judge in the Bits of Freedom vs. Meta lawsuit: Meta must respect users' choice*. Accessed: 2025-10-22. URL: https://www.bitsoffreedom.nl/2025/10/02/judge-in-the-bits-of-freedom-vs-meta-lawsuit-meta-must-respect-users-choice/.
— (2026). *Court again rules in favor of Bits of Freedom: freedom of choice for Instagram and Facebook users remains intact*. URL: https://www.bitsoffreedom.nl/2026/03/10/court-again-rules-in-favor-of-bits-of-freedom-freedom-of-choice-for-instagram-and-facebook-users-remains-intact/.
Bodo, Balazs (2025). "Humble tools of divine intervention–The misunderstood role of algorithms in public opinion formation". In: *Dialogues on Digital Society*.

Bonicalzi, Sofia, Mario De Caro, and Benedetta Giovanola (2023). "Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems". In: *Topoi* 42.3, pp. 819–832. DOI: 10.1007/s11245-023-09922-5.

Bostandjiev, Svetlin, John O'Donovan, and Tobias Höllerer (2012a). "TasteWeights: a visual interactive hybrid recommender system". In: *Proceedings of the Sixth ACM Conference on Recommender Systems*. RecSys '12. Dublin, Ireland: Association for Computing Machinery, pp. 35–42. ISBN: 9781450312707.

— (2012b). "TasteWeights: a visual interactive hybrid recommender system". In: *Sixth ACM Conference on Recommender Systems, RecSys '12, Dublin, Ireland, September 9-13, 2012*. Ed. by Padraig Cunningham et al. ACM, pp. 35–42.

Bozdag, Engin (2013). "Bias in algorithmic filtering and personalization". In: *Ethics and information technology* 15, pp. 209–227.

Bozdag, Engin and Jeroen Van Den Hoven (2015). "Breaking the filter bubble: democracy and design". In: *Ethics and information technology* 17, pp. 249–265.

Breton, Thierry (2024). *Be our guest @elonmusk*. Accessed on 16 December 2024. URL: https://x.com/ThierryBreton/status/1811811489889

Bullock, Justin B (2019). "Artificial intelligence, discretion, and bureaucracy". In: *The American Review of Public Administration* 49.7, pp. 751–761.

Busch, Christoph (2023). "From algorithmic transparency to algorithmic choice: European perspectives on recommender Systems and platform regulation". In: *Recommender Systems: Legal and Ethical Issues*. Springer International Publishing Cham, pp. 31–54.

Carter, Ian (2022). "Positive and Negative Liberty". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2022. Metaphysics Research Lab, Stanford University.

Carvalho, Mateus Correia de and Rachel Griffin (2026). "Who speaks and who is heard? Civil society participation and participatory justice in DSA systemic risk management". In: *DSA Observatory*. URL: https://dsa-observatory.eu/2026/02/17/who-speaks-and-who-is-heard/.

Cena, Federica et al. (2017). "How scales influence user rating behaviour in recommender systems". In: *Behaviour & Information Technology* 36.10, pp. 985–1004.

Centre for Democracy and Technology (2023). *CDT and Technologists File SCOTUS Brief Urging Court To Hold that Section 230 Applies to Recommendations of Content*. Accessed 05-02-2023. URL: https://cdt.

org/insights/cdt-and-technologists-file-scotus-brief-urging-court-to-hold-that-section-230-applies-to-recommendations-of-content/.
Chernev, Alexander (Sept. 2003). "When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice". In: *Journal of Consumer Research* 30.2, pp. 170–183. ISSN: 0093-5301.
Clark, Emma (Feb. 2023). *Elon Musk created a special system for showing you all his tweets first*. The Verge. Accessed: 2025-11-03. URL: https://www.theverge.com/2023/2/14/23600358/elon-musk-tweets-algorithm-changes-twitter.
Commission, European (2024a). *Artificial Intelligence Act enters into force*. Accessed: 2025-10-22. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4123.
— (2024b). *Commission opens formal proceedings against TikTok on election risks under the Digital Services Act*. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_6487.
— (2024c). *Commission opens formal proceedings against TikTok on election risks under the Digital Services Act*. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_6487.
— (2024d). *DSA.100121 - TikTok - INVESTIGATION INTO COMPLIANCE WITH ARTICLES 34, 35 AND 42(4)*. URL: https://ec.europa.eu/newsroom/dae/redirection/document/108626.
— (2025). *Guidelines on measures to ensure a high level of privacy, safety and security for minors online, pursuant to Article 28(4) of Regulation (EU) 2022/2065*. URL: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-protection-minors.
Conati, Cristina et al. (2014). "Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making". In: *Computer Graphics Forum*. Vol. 33. 3. Wiley Online Library, pp. 371–380.
Costello, Nancy et al. (2023). "Algorithms, addiction, and adolescent mental health: An interdisciplinary study to inform state-level policy action to protect youth from the dangers of social media". In: *American Journal of Law & Medicine* 49.2-3, pp. 135–172.
Covington, Paul, Jay Adams, and Emre Sargin (2016). "Deep neural networks for youtube recommendations". In: *Proceedings of the 10th ACM conference on recommender systems*, pp. 191–198.

Deley, Trevor and Elizabeth Dubois (2020). "Assessing trust versus reliance for technology platforms by systematic literature review". In: *Social Media+ Society* 6.2, p. 2056305120913883.

Desrosiers, Christian and George Karypis (Jan. 2011). "A Comprehensive Survey of Neighborhood-Based Recommendation Methods". In: *Recommender Systems Handbook*, pp. 107–144. ISBN: 978-0-387-85819-7.

District Court of Amsterdam (Oct. 2, 2025). *Judgment in preliminary relief proceedings of 2 October 2025: Stichting Bits of Freedom vs. Facebook Netherlands B.V. & Meta Platforms Ireland Ltd.* Accessed: 2025-10-22. URL: https://www.bitsoffreedom.nl/wp-content/uploads/2025/10/20251002-District-Court-of-Amsterdam_en.pdf.

Dokoupil, Patrik, Ludovico Boratto, and Ladislav Peska (2024a). "Comparing User Interfaces for Customizing Multi-Objective Recommender Systems". In: *Proceedings of the 11th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems*. IntRS@RecSys'24. Bari, Italy: CEUR-WS.org, pp. 15–28. URL: https://ceur-ws.org/Vol-3815/paper2.pdf.

— (2024b). "User Perceptions of Diversity in Recommender Systems". In: *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2024, Cagliari, Italy, July 1-4, 2024*. ACM, pp. 212–222.

Dokoupil, Patrik, Ladislav Peska, and Ludovico Boratto (2023). "Looks Can Be Deceiving: Linking User-Item Interactions and User's Propensity Towards Multi-Objective Recommendations". In: *Proceedings of the 17th ACM Conference on Recommender Systems*. RecSys '23. Singapore, Singapore: Association for Computing Machinery, pp. 912–918. ISBN: 9798400702419.

Dooms, Simon, Toon De Pessemier, and Luc Martens (2014). "Improving IMDb Movie Recommendations with Interactive Settings and Filters". In: *Poster Proceedings of the 8th ACM Conference on Recommender Systems, RecSys 2014, Foster City, Silicon Valley, CA, USA, October 6-10, 2014*. Ed. by Li Chen and Jalal Mahmud. Vol. 1247. CEUR Workshop Proceedings. CEUR-WS.org. URL: https://ceur-ws.org/Vol-1247/recsys14%5C_poster19.pdf.

Edwards, Jonathan (2022). *Mother sues TikTok after 10-year-old died trying 'Blackout Challenge'*. Accessed 11-08-2022. URL: https://www.washingtonpost.com/nation/2022/05/17/tiktok-blackout-challenge-lawsuit/.

Egger, Roman and Joanne Yu (May 2022). "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts". In: *Front. Sociol.* 7. ISSN: 2297-7775.

El Majjodi, Ayoub, Alain D Starke, and Christoph Trattner (2022). "Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system". In: *Proceedings of the 30th ACM conference on user modeling, adaptation and personalization*, pp. 48–56.

Epstein, Ziv, Gordon Pennycook, and David Rand (2020). "Will the crowd game the algorithm? Using layperson judgments to combat misinformation on social media by downranking distrusted sources". In: *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–11.

Eslami, Motahhare et al. (2015). "" I always assumed that I wasn't really that close to [her]" Reasoning about Invisible Algorithms in News Feeds". In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pp. 153–162.

European Centre for Algorithmic Transparency (n.d.[a]). *About ECAT*. Accessed on 10 May 2024. URL: https://algorithmic-transparency.ec.europa.eu/about_en.

— (n.d.). *About ECAT*. URL: https://algorithmic-transparency.ec.europa.eu/about_en.

— (n.d.[b]). *Who we are*. Accessed on 10 May 2024. URL: https://algorithmic-transparency.ec.europa.eu/team_en.

European Commission (2016). *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Accessed 20-08-2022. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN.

— (2020). *Guidelines on ranking transparency pursuant to Regulation (EU) 2019/1150 of the European Parliament and of the Council*. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020XC1208(01).

— (2023a). *Commission decision initiating proceedings pursuant to Article 66(1) of Regulation (EU) 2022/2065*. Accessed on 12 December 2024. URL: https://ec.europa.eu/newsroom/dae/redirection/document/101292.

— (2023b). *Commission delegated regulation supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council, by laying down rules on the performance of audits for very large online platforms and very large online search engines*. URL: https://ec.europa.eu/newsroom/dae/redirection/document/99436.
— (2023c). *Commission opens formal proceedings against X under the Digital Services Act*. Accessed on 12 December 2024. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709.
— (2023d). *The commission sends request for information to X under the Digital Services Act*. Accessed on 12 December 2024. URL: https://ec.europa.eu/commission/presscorner/detail/en/IP_23_4953.
— (2024a). *Call for tenders EC-CNECT/2024/OP/0032 - Digital Services Act: Study on systemic risks and their mitigation*. URL: %5Curl%7Bhttps://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/tender-details/c7f44485-4fac-457a-95ca-4a1671ac7b8d-CN?isExactMatch=true%5C&order=DESC%5C&pageNumber=1%5C&pageSize=50%5C&sortBy=startDate%7D (visited on 12/16/2024).
— (2024b). *Call for tenders EC-CNECT/2024/OP/0052 - Digital Services Act: Technical assistance for market intelligence, evidence gathering and compliance monitoring*. URL: %5Curl%7Bhttps://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/tender-details/efd2992a-c0a2-498c-83ca-a1729a3863ff-CN%7D (visited on 12/16/2024).
— (2024c). *Commission opens formal proceedings against TikTok on election risks under the Digital Services Act*. URL: %5Curl%7Bhttps://ec.europa.eu/commission/presscorner/detail/en/ip_24_6487%7D (visited on 12/16/2024).
— (2024d). *Commission opens formal proceedings against TikTok under the Digital Services Act*. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926.
— (2024e). *Commission sends preliminary findings to X for breach of the Digital Services Act*. Accessed on 12 December 2024. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761.
— (2024f). *Commission sends requests for information to YouTube, Snapchat, and TikTok on recommender systems under the Digital Services Act*. URL: https://digital-strategy.ec.europa.eu/en/news/

commission-sends-requests-information-youtube-snapchat-and-tiktok-recommender-systems-under-digital.
European Commission (2024g). *Commission sends requests for information to YouTube, Snapchat, and TikTok on recommender systems under the Digital Services Act*. URL: %5Curl%7Bhttps://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-youtube-snapchat-and-tiktok-recommender-systems-under-digital%7D (visited on 12/16/2024).
— (2024h). *Do you want to help enforce the Digital Services Act? Apply now to be part of the DSA enforcement team!* Accessed on 12 December 2024. URL: https://digital-strategy.ec.europa.eu/en/news/do-you-want-help-enforce-digital-services-act-apply-now-be-part-dsa-enforcement-team.
— (2024i). *European Board for Digital Services*. Accessed on 16 December 2024. URL: https://digital-strategy.ec.europa.eu/en/policies/dsa-board.
— (2024j). *Job opportunities within the Digital Services Act enforcement team*. Accessed on 12 December 2024. URL: https://eu-careers.europa.eu/sites/default/files/eu_vacancies/2024-01/2024_2nd%20call_Job_opportunities%20within%20the%20DSA%20Enforcement%20Team%20final%20with%20privacy%20statement%20Final.pdf.
— (2024k). *Supervision of the designated very large online platforms and search engines under DSA*. Accessed on 16 December 2024. URL: https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses.
— (2025a). *Call for tenders EC-CNECT/2025/OP/0056 - User observations on usage of recommender systems to enable scaling of recommender systems experiments through automated and agentic means*. Accessed: 2025-10-22. URL: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/tender-details/docs/4ab71ff5-0048-44ab-ac98-c5bb2c7f527b-CN/EN_TENDER_SPECIFICATIONS_EC-CNECT-2025-OP-0056_User%20observations%20Recommender%20Systems_V1.pdf.
— (2025b). *Code of Conduct on Disinformation*. URL: https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation.
— (2025c). *COMMISSION OPINION of 13.2.2025 on the assessment of the Code of Practice on Disinformation within the meaning of Article 45 of Reg-*

ulation 2022/2065. URL: https://ec.europa.eu/newsroom/dae/redirection/document/112679.
— (2025d). *Commission preliminarily finds TikTok and Meta in breach of their transparency obligations under the Digital Services Act*. Accessed: 2025-10-25. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2503.
— (2025e). *Delegated act on data access under the Digital Services Act (DSA)*. URL: https://digital-strategy.ec.europa.eu/en/library/delegated-act-data-access-under-digital-services-act-dsa.
— (2025f). *Guidelines on measures to ensure a high level of privacy, safety and security for minors online, pursuant to Article 28(4) of Regulation (EU) 2022/2065*. URL: https://ec.europa.eu/newsroom/dae/redirection/document/118226.
— (2025g). *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus)*. URL: https://eur-lex.europa.eu/resource.html?uri=cellar:ebf17714-c56e-11f0-8da2-01aa75ed71a1.0001.02/DOC_1&format=PDF.
— (2025h). *The Code of Conduct on Disinformation*. URL: https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation.
— (n.d.). *HUMAINT publications*. Accessed on 20 February 2025. URL: https://ai-watch.ec.europa.eu/humaint/publications_en.
European Ombudsman (2024). *The European Ombudsman's preliminary views on the European Commission's refusal to give public access to the risk assessment report of a large social media company on its compliance with the provisions of the Digital Services Act*. Accessed on 20 February 2025. URL: https://www.ombudsman.europa.eu/en/doc/correspondence/en/199731.
European Parliament and Council (2001). *REGULATION (EC) No 1049/2001 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 30 May 2001 regarding public access to European Parliament, Council and Commission documents*. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32001R1049.

European Parliament and Council (2019). *REGULATION (EU) 2019/1150 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services*. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R1150.
— (2022). *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)*. URL: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng.
— (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.
— (n.d.). *Regulation (EU) 2024/1083 of the European Parliament and of the Council of 11 April 2024 establishing a common framework for media services in the internal market and amending Directive 2010/13/EU (European Media Freedom Act)*. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1083.
EY (2024a). *Independent Audit on Instagram*. URL: https://transparency.meta.com/sr/dsa-independent_audit_report_Aug_2023-Jun_2024_instagram.
— (2024b). *Independent Audit on YouTube*. URL: https://storage.googleapis.com/transparencyreport/report-downloads/dsa-audit-youtube_2023-8-28_2024-5-31_en_v1.pdf.
Fabbri, Matteo (2023a). "How and to which extent will the provisions of the Digital Services Act of the European Union impact on the relationship between users and platforms as information providers?" In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 1002–1003.
— (2023b). "Self-determination through explanation: an ethical perspective on the implementation of the transparency requirements for recommender systems set by the Digital Services Act of the European Union". In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 653–661.
— (2023c). "Social influence for societal interest: a pro-ethical framework for improving human decision making through multi-stakeholder recommender systems". In: *AI & SOCIETY* 38.2, pp. 995–1002.

— (2025a). "A Classification Framework for Algorithmic Recommendations: System, Agent, Patient". In: *Frontiers in Artificial Intelligence and Applications, Volume 408: HHAI 2025*. Amsterdam, The Netherlands: IOS Press, pp. 449–454.
— (2025b). "The Role of Requests for Information in Governing Digital Platforms Under the Digital Services Act: The Case of X". In: *Journalism and Media* 6.1, p. 41.
— (2026). "What the EU's X Decision Reveals About How the DSA Is Enforced". In: *Tech Policy Press*. URL: https://www.techpolicy.press/what-the-eus-x-decision-reveals-about-how-the-dsa-is-enforced/.
Fabbri, Matteo and Ludovico Boratto (2025). "Auditing Recommender Systems for User Empowerment in Very Large Online Platforms under the Digital Services Act". In: *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, pp. 51–61.
Field, M. B. and S. Roberts (2020). "Trust, integrity and the weaponising of information: The EU's transparency paradox". In: *Journal of Contemporary European Research* 16.3.
Floridi, Luciano (2008). "The method of levels of abstraction". In: *Minds and machines* 18.3, pp. 303–329.
— (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.
— (2015a). "The new grey power". In: *Philosophy & Technology* 28, pp. 329–332.
— (2015b). *The onlife manifesto: Being human in a hyperconnected era*. Springer nature.
Gabeur, Valentin et al. (2020). "Multi-modal transformer for video retrieval". In: *European Conference on Computer Vision*. Springer, pp. 214–229.
Gandini, Alessandro, Silvia Keeling, and Urbano Reviglio (2025). "Conceptualising the 'algorithmic public opinion': Public opinion formation in the digital age". In: *Dialogues on Digital Society*, p. 29768640251323147.
Gedikli, Fatih, Dietmar Jannach, and Mouzhi Ge (2014). "How should I explain? A comparison of different explanation types for recommender systems". In: *International Journal of Human-Computer Studies* 72.4, pp. 367–382. ISSN: 1071-5819.
Gillespie, Tarleton (2014). "The relevance of algorithms". In: *Media technologies: Essays on communication, materiality, and society* 167.2014, p. 167.

Gillingham, Cara, Jessica Morley, and Luciano Floridi (2025). "The Effects of AI on Street-Level Bureaucracy: A Scoping Review". In: *Digital Society* 4.1, p. 22.

Google (2024a). *Code of Practice of Disinformation – Report of Google for the period of 1 January 2024 to 30 June 2024*. URL: https://disinfocode.eu/reports-archive/reports-september-2024/.

— (2024b). *Report of Systemic Risk Assessments*. URL: https://storage.googleapis.com/transparencyreport/report-downloads/dsa-risk-assessment%5C_2023-8-28%5C_2023-8-28%5C_en%5C_v1.pdf.

Google Ireland Limited (2024). *DSA Audit Implementation Report*. URL: https://storage.googleapis.com/transparencyreport/report-downloads/dsa-audit-google-implementation%5C_2023-8-28%5C_2024-5-31%5C_en%5C_v1.pdf.

Gorgoglione, Michele, Umberto Panniello, and Alexander Tuzhilin (2019). "Recommendation strategies in personalization applications". In: *Information & Management* 56.6, p. 103143.

Grisse, Karina (2023). "Recommender Systems, Manipulation and Private Autonomy: How European Civil Law Regulates and Should Regulate Recommender Systems for the Benefit of Private Autonomy". In: *Recommender Systems: Legal and Ethical Issues*. Ed. by Sergio Genovesi, Katharina Kaesling, and Scott Robbins. Cham: Springer International Publishing, pp. 101–128. ISBN: 978-3-031-34804-4.

Grootendorst, Maarten (Mar. 2022). *BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure*. arXiv: 2203.05794 [cs].

Hampton, Lelia Marie (2021). "Black feminist musings on algorithmic oppression". In: *arXiv preprint arXiv:2101.09869*.

Harambam, Jaron, Natali Helberger, and Joris Van Hoboken (2018). "Democratizing algorithmic news recommenders: how to materialize voice in a technologically saturated media ecosystem". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2133, p. 20180088.

Harper, F. Maxwell et al. (2015). "Putting Users in Control of their Recommendations". In: *Proceedings of the 9th ACM Conference on Recommender Systems*. RecSys '15. Vienna, Austria: Association for Computing Machinery, pp. 3–10. ISBN: 9781450336925.

Helberger, Natali, Jo Pierson, and Thomas Poell (2018). "Governing online platforms: From contested to cooperative responsibility". In: *The information society* 34.1, pp. 1–14.

Helberger, Natali, Marijn Sax, et al. (2022). "Choice architectures in the digital economy: Towards a new understanding of digital vulnerability". In: *Journal of Consumer Policy* 45.2, pp. 175–200.

Helberger, Natali, Max Van Drunen, et al. (2021). "Regulation of news recommenders in the Digital Services Act: Empowering David against the very large online Goliath". In: *Internet Policy Review* 26.

Help, YouTube (2025). *Manage your recommendations & search results*. https://support.google.com/youtube/answer/6342839?hl=en. [Accessed 13-07-2025].

Jahangir, Ramsha (2025). "What a Dutch Court Ruling Against Meta Signals for Private DSA Enforcement". In: *Tech Policy Press*. Accessed: 2025-10-22. URL: https://www.techpolicy.press/what-a-dutch-court-ruling-against-meta-signals-for-private-dsa-enforcement/.

Jahangir, Ramsha and Justin Hendrix (2026). *EU Decision Behind €120m Fine on Musk's X Released by US Lawmakers*. URL: https://www.techpolicy.press/eu-decision-behind-120m-fine-on-musks-x-released-by-us-lawmakers/.

Jannach, Dietmar et al. (2015). "What recommenders recommend: an analysis of recommendation biases and possible countermeasures". In: *User Modeling and User-Adapted Interaction* 25, pp. 427–491.

Jelodar, Hamed et al. (Dec. 2018). *Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey*. arXiv: 1711.04305 [cs].

Jesse, Mathias and Dietmar Jannach (Jan. 2021). "Digital Nudging with Recommender Systems: Survey and Future Directions". In: *Computers in Human Behavior Reports* 3. ISSN: 2451-9588.

Jin, Yucheng, Bruno De Lemos Ribeiro Pinto Cardoso, and Katrien Verbert (2017). "How do different levels of user control affect cognitive load and acceptance of recommendations?" In: *IntRS@ RecSys*, pp. 35–42.

Jin, Yucheng, Nava Tintarev, et al. (2020). "Effects of personal characteristics in control-oriented user interfaces for music recommender systems". In: *User Modeling and User-Adapted Interaction* 30.2, pp. 199–249.

Kant, Immanuel (2013). *An answer to the question:'What is enlightenment?'* Penguin UK.

Kherwa, Pooja and Poonam Bansal (July 2018). "Topic Modeling: A Comprehensive Review". In: *ICST Transactions on Scalable Information Systems* 7, p. 159623.

Kluttz, Daniel N and Deirdre K Mulligan (2019). "Automated decision support technologies and the legal profession". In: *Berkeley Technology Law Journal* 34.3, pp. 853–890.

Knijnenburg, Bart P et al. (2012). "Inspectability and control in social recommenders". In: *Proceedings of the sixth ACM conference on Recommender systems*, pp. 43–50.

— (2012). "Inspectability and control in social recommenders". In: *Proceedings of the Sixth ACM Conference on Recommender Systems*. RecSys '12. Dublin, Ireland: Association for Computing Machinery, pp. 43–50. ISBN: 9781450312707.

Knijnenburg, Bart P., Niels J.M. Reijmer, and Martijn C. Willemsen (2011). "Each to His Own: How Different Users Call for Different Interaction Methods in Recommender Systems". In: *Proceedings of the Fifth ACM Conference on Recommender Systems*. RecSys '11. Chicago, Illinois, USA: Association for Computing Machinery, pp. 141–148. ISBN: 9781450306836. DOI: 10.1145/2043932.2043960.

Kolkman, Daan (2020). *"F**k the algorithm"?: What the world can learn from the UK's A-level grading fiasco*. LSE Impact Blog. Accessed: 2025-10-22. URL: https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/.

Kostopoulos, Georgios, Gregory Davrazos, and Sotiris Kotsiantis (2024). "Explainable Artificial Intelligence-Based Decision Support Systems: A Recent Review". In: *Electronics* 13.14. ISSN: 2079-9292.

KPMG (2024). *DSA Assurance Report*. URL: https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy%5B/ljhwZthlaukjlkulzlp/DSA_H2_2024/TikTok-DSA-Audit-Assurance-Report-Sep-2024.pdf.

Lagioia Francesca, Giovanni Sartor and Andrea Simoncini (2018). *Commento all'articolo 22 del GDPR*.

Leerssen, Paddy (2020). "The soap box as a black box: Regulating transparency in social media recommender systems". In: *European Journal of Law and Technology* 11.2.

— (2023). "An end to shadow banning? Transparency rights in the Digital Services Act between content moderation and curation". In: *Computer Law & Security Review* 48.

Levenson, Michael and April Rubin (2022). *Parents Sue TikTok, Saying Children Died After Viewing 'Blackout Challenge'*. Accessed 15-01-2023. URL: https://www.nytimes.com/2022/07/06/technology/tiktok-blackout-challenge-deaths.html.

Liang, Yu, Aditya Ponnada, et al. (2023). "Enabling Goal-Focused Exploration of Podcasts in Interactive Recommender Systems". In: *Proceedings of the 28th International Conference on Intelligent User Interfaces*. IUI '23. Sydney, NSW, Australia: Association for Computing Machinery, pp. 142–155. ISBN: 9798400701061.

Liang, Yu and Martijn C. Willemsen (2022a). "Exploring the longitudinal effects of nudging on users' music genre exploration behavior and listening preferences". In: *RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022*. Ed. by Jennifer Golbeck et al. ACM, pp. 3–13. DOI: 10.1145/3523227.3546772. URL: https://doi.org/10.1145/3523227.3546772.

— (2022b). "Exploring the longitudinal effects of nudging on users' music genre exploration behavior and listening preferences". In: *Proceedings of the 16th ACM Conference on Recommender Systems*. RecSys '22. Seattle, WA, USA: Association for Computing Machinery, pp. 3–13. ISBN: 9781450392785.

Liesenfeld, Anna (2025). "The legal significance of independent research based on article 40 DSA for the management of systemic risks in the digital services act". In: *European Journal of Risk Regulation* 16.1, pp. 184–196.

Light, Ben, Jean Burgess, and Stefanie Duguay (2018). "The walkthrough method: An approach to the study of apps". In: *New media & society* 20.3, pp. 881–900.

Loi, Michele, Matteo Fabbri, and Andrea Ferrario (2025). "Regulating the Undefined: Addressing Systemic Risks in the Digital Services Act (with an Appendix on the AI Act)". In: *Philosophy & Technology* 38.2, p. 82.

Madhani, Aamer (2025). *Emerging TikTok deal with China ensures US control of board and crucial algorithm, White House says*. Associated Press. Accessed: 2025-10-22. URL: https://apnews.com/article/tiktok-china-trump-social-video-platform-oracle-cbffc62506b3f4533ed327a37ac38ad3.

Marciano, Laura et al. (2025). "Digital Media, Cognition, and Brain Development in Adolescence". In: *Handbook of Children and Screens*. Springer, Cham, pp. 21–29.

Mathur, Arunesh, Mihir Kshirsagar, and Jonathan Mayer (2021). "What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods". In: *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–18.

McInerney, James et al. (2018). "Explore, exploit, and explain: personalizing explainable recommendations with bandits". In: *Proceedings of the 12th ACM conference on recommender systems*, pp. 31–39.
McNee, Sean M. et al. (2003). "Interfaces for Eliciting New User Preferences in Recommender Systems". In: *User Modeling 2003*. Ed. by Peter Brusilovsky, Albert Corbett, and Fiorella de Rosis. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 178–187. ISBN: 978-3-540-44963-8.
Meta (2024a). *Audit Implementation Report for Instagram Regarding Regulation (EU) 2022/2065, the Digital Services Act (DSA)*. URL: https://transparency.meta.com/sr/dsa-audit_implementation_report_Sept_2024_instagram.
— (2024b). *Our approach to explaining ranking*. URL: https://transparency.meta.com/features/explaining-ranking/.
— (2024c). *Regulation (EU) 2022/2065 Digital Services Act (DSA) - Systemic Risk Assessment and Mitigation Report for Instagram*. URL: https://scontent-dus1-1.xx.fbcdn.net/v/t39.8562-6/468481711_1994741217693263_5869766641031282981_n.pdf?_nc_cat=106&ccb=1-7&_nc_sid=b8d81d&_nc_ohc=A0tw9NWkjwkQ7kNvgGkhfWJ&_nc_zt=14&_nc_ht=scontent-dus1-1.xx&_nc_gid=ADiAYO2x9S8umqamoVAcgU-&oh=00_AYDsWtjJiy1_8koPrkj8eywHJVTPGd3VMWNNdr-GZpMdwg&oe=67970FF5.
— (2025). *More Speech and Fewer Mistakes*. URL: https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/ (visited on 02/16/2025).
— (n.d.). *Facebook preliminary views and comments on the Digital Services Act*. Accessed 20-07-2022. URL: https://enterprise.gov.ie/en/consultations/consultations-files/facebook-dsa-submission.pdf.
Milano, Silvia, Brent Mittelstadt, et al. (2021). "Epistemic fragmentation poses a threat to the governance of online targeting". In: *Nature Machine Intelligence* 3.6, pp. 466–472.
Milano, Silvia, Mariarosaria Taddeo, and Luciano Floridi (2020). "Recommender systems and their ethical challenges". In: *Ai & Society* 35, pp. 957–967.
— (2021). "Ethical aspects of multi-stakeholder recommendation systems". In: *The information society* 37.1, pp. 35–45.
Millecamp, Martijn et al. (2018). "Controlling Spotify Recommendations: Effects of Personal Characteristics on Music Recommender User Interfaces". In: *Proceedings of the 26th Conference on User Modeling, Adap-*

tation and Personalization. UMAP '18. Singapore, Singapore: Association for Computing Machinery, pp. 101–109. ISBN: 9781450355896.

Moens, Barbara (2025). "EU fines Apple and Meta total of €700 mn for antitrust violations". In: *Financial Times*. URL: https://www.ft.com/content/09af8a25-6cf9-4bf1-8cd8-16884416e27f (visited on 11/15/2025).

Montag, Christian et al. (2019). "Addictive features of social media/messenger platforms and freemium games against the background of psychological and economic theories". In: *International journal of environmental research and public health* 16.14, p. 2612.

Nechushtai, Efrat and Seth C Lewis (2019). "What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations". In: *Computers in human behavior* 90, pp. 298–307.

Noble, Safiya Umoja (2018). "Algorithms of oppression: How search engines reinforce racism". In: *Algorithms of oppression*. New York university press.

O'Donovan, John et al. (2008). "PeerChooser: visual interactive recommendation". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. Florence, Italy: Association for Computing Machinery, pp. 1085–1088. ISBN: 9781605580111.

Panoptykon Foundation, Irish Council for Civil Liberties, and People vs Big Tech (2023). *Fixing Recommender Systems: From identification of risk factors to meaningful transparency and mitigation*. URL: https://panoptykon.org/sites/default/files/2023-08/Panoptykon_ICCL_PvsBT_Fixing-recommender-systems_Aug%202023.pdf.

Panoptykon Foundation and People vs Big Tech (2023). *Prototyping User Empowerment: Towards DSA-Compliant Recommender Systems*. URL: https://panoptykon.org/sites/default/files/2023-11/peoplevsbigtec panoptykon_prototyping-empowerment_brief_20112023.pdf (visited on 11/15/2025).

Parliament, European (2023). *Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*. Accessed 06-07-2023. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html.

Parra, Denis and Peter Brusilovsky (2015). "User-controllable personalization: A case study with SetFusion". In: *International Journal of Human-Computer Studies* 78, pp. 43–67. ISSN: 1071-5819.

Parra, Denis, Peter Brusilovsky, and Christoph Trattner (2014). "See what you want to see: visual user-driven approach for hybrid recommendation". In: *19th International Conference on Intelligent User Interfaces, IUI 2014, Haifa, Israel, February 24-27, 2014*. Ed. by Tsvi Kuflik et al. ACM, pp. 235–240.

Pedreschi, Dino et al. (2025). "Human-AI coevolution". In: *Artificial Intelligence* 339, p. 104244.

Qin, Yao, Bahiyah Omar, and Alessandro Musetti (2022). "The addiction behavior of short-form video app TikTok: The information quality and system quality perspective". In: *Frontiers in Psychology* 13, p. 932805.

Radford, Alec et al. (2021). "Learning Transferable Visual Models From Natural Language Supervision". In: *Proceedings of the 38th International Conference on Machine Learning, ICML 2021*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. Cambridge, MA: PMLR, pp. 8748–8763.

Rani, Neha, Sharon Lynn Chu, and Victoria Rene Mei (2022). "Investigating the Effects of Different Levels of User Control on the Effectiveness of Context-Aware Recommender Systems for Web-Based Search". In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI EA '22. New Orleans, LA, USA: Association for Computing Machinery. ISBN: 9781450391566.

Reuters (2025). *Dutch court orders Meta to change Facebook and Instagram timeline settings*. Accessed: 2025-10-22. URL: https://www.reuters.com/technology/dutch-court-orders-meta-change-facebook-instagram-timeline-settings-2025-10-02/.

Reviglio, Urbano and Matteo Fabbri (2024). "Navigating the Digital Services Act: Scenarios of Transparency and User Control in VLOPs' Recommender Systems". In: *NORMalize 2024: The Second Workshop on the Normative Design and Evaluation of Recommender Systems, October 18, 2024, co-located with the ACM Conference on Recommender Systems 2024 (RecSys 2024), Bari, Italy*.

Reviglio, Urbano and Giulio Santoni (2023). "Governing Platform Recommender Systems in Europe: Insights from China". In: *Global Jurist* 23.2, pp. 151–181.

Ricci, Francesco et al. (2022). *Recommender Systems Handbook*. Ed. by Bracha Shapira Francesco Ricci Lior Rokach. 3rd. Berlin, Heidelberg: Springer New York, NY. ISBN: 978-1-07-162196-7.

Ryan-Mosley, Tate (2023). *How the Supreme Court ruling on Section 230 could end Reddit as we know it*. Accessed 05-02-2023. URL: https://www.technologyreview.com/2023/02/01/1067520/supreme-court-section-230-gonzalez-reddit/.

Sala, Arianna, Lorenzo Porcaro, and Emilia Gómez (2024). "Social Media Use and adolescents' mental health and well-being: An umbrella review". In: *Computers in Human Behavior Reports* 14, p. 100404.

Sarkar, Alisha Rahaman (2022). *TikTok's 'blackout' challenge linked to deaths of 20 children in 18 months, report says*. Accessed 15-01-2023. URL: https://www.independent.co.uk/tech/tiktok-blackout-challenge-deaths-b2236669.html.

Sax, Marijn (2021). *Between empowerment and manipulation*. Wolters Kluwer.

Schaffer, James, Tobias Höllerer, and John O'Donovan (2015). "Hypothetical Recommendation: A Study of Interactive Profile Manipulation Behavior for Recommender Systems." In: *FLAIRS*. Vol. 15, pp. 507–512.

Shepardson, David and Alexandra Alper (2025). "US lawmaker says licensing deal for TikTok algorithm would raise serious concerns". In: *Reuters*. Accessed: 2025-10-22. URL: https://www.reuters.com/world/us-lawmaker-says-licensing-deal-tiktok-algorithm-would-raise-serious-concerns-2025-10-16/.

Shuyo, Nakatami (n.d.). *langdetect 1.0.9*. URL: https://pypi.org/project/langdetect/.

Söderlund, Kasia et al. (2024). "Regulating high-reach AI: On transparency directions in the Digital Services Act". In: *Internet Policy Review* 13.1, pp. 1–31.

Starke, Christopher et al. (2025a). "Contesting personalized recommender systems: a cross-country analysis of user preferences". In: *Information, Communication & Society* 28.1, pp. 41–60.

— (2025b). "Contesting personalized recommender systems: a cross-country analysis of user preferences". In: *Information, Communication & Society* 28.1, pp. 41–60.

Stray, Jonathan et al. (Nov. 2023). "Building Human Values into Recommender Systems: An Interdisciplinary Synthesis". In: *ACM Trans. Recomm. Syst.*

Sun, Ruixuan et al. (2023). "Interactive Content Diversity and User Exploration in Online Movie Recommenders: A Field Experiment". In: *International Journal of Human–Computer Interaction* 0.0, pp. 1–15.
— (2024). "Interactive Content Diversity and User Exploration in Online Movie Recommenders: A Field Experiment". In: *Int. J. Hum. Comput. Interact.* 40.22, pp. 7233–7247.
Sunstein, Cass (2018). # *Republic: Divided democracy in the age of social media*. Princeton University Press.
Supreme Court of the United States (2022). *21-1333 GONZALEZ V. GOOGLE LLC*. Accessed 05-02-2023. URL: https://www.supremecourt.gov/qp/21-01333qp.pdf.
— (2023). *21-1333 REYNALDO GONZALEZ, ET AL., PETITIONERS v. GOOGLE LLC ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT*. Accessed 25-06-2023. URL: https://www.supremecourt.gov/opinions/22pdf/21-1333_6j7a.pdf.
Terzis, Petros, Michael Veale, and Noëlle Gaumann (2024). "Law and the Emerging Political Economy of Algorithmic Audits". In: *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1255–1267.
Thaler, Richard and Cass Sunstein (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.
TikTok (2024a). *DSA Audit Implementation Report*. URL: https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy%5B/ljhwZthlaukjlkulzlp/DSA_H2_2024/TikTok-DSA-Audit-Implementation-Report-Sep-2024.pdf.
— (2024b). *DSA RISK ASSESSMENT REPORT 2023*. URL: https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy%5B/ljhwZthlaukjlkulzlp/DSA_H2_2024/TikTok-DSA-Risk-Assessment-Report-2023.pdf.
— (2024c). *TikTok's STEM feed now available to all users by default*. URL: https://newsroom.tiktok.com/en-gb/tiktok-stem-feed-now-available-to-all-users-by-default.
— (2025). *DSA RISK ASSESSMENT REPORT 2025*. URL: https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy%5B/ljhwZthlaukjlkulzlp/DSA/DSARiskAssessmentReport-TikTokTechnologyLimited-2025.pdf.
— (n.d.). *How TikTok recommends content*. URL: https://support.tiktok.com/en/using-tiktok/exploring-videos/how-tiktok-recommends-content.

Tintarev, Nava and Judith Masthoff (2007). "A survey of explanations in recommender systems". In: *2007 IEEE 23rd international conference on data engineering workshop*. IEEE, pp. 801–810.

— (2012). "Evaluating the effectiveness of explanations for recommender systems: Methodological issues and empirical studies on the impact of personalization". In: *User Modeling and User-Adapted Interaction* 22, pp. 399–439.

— (2015). "Explaining recommendations: Design and evaluation". In: *Recommender systems handbook*. Springer, pp. 353–382.

Tkachenko, Yegor, Wassim Dhaouadi, and Kamel Jedidi (2022). "Scaling up Ranking under Constraints for Live Recommendations by Replacing Optimization with Prediction". In: arXiv: 2202.07088.

Tsai, Chun-Hua and Peter Brusilovsky (2021). "The effects of controllability and explainability in a social recommender system". In: *User Modeling and User-Adapted Interaction* 31, pp. 591–627.

Vincent, Cindy S and Adam Gismondi (2021). "Fake news, reality apathy, and the erosion of trust and authenticity in American politics". In: *Democracy in the Disinformation Age*. Routledge, pp. 79–98.

Violot, Caroline et al. (2024). "Shorts vs. Regular Videos on YouTube: A Comparative Analysis of User Engagement and Content Creation Trends". In: *Proceedings of the 16th ACM Web Science Conference*. WEBSCI '24. Stuttgart, Germany: Association for Computing Machinery, pp. 213–223. ISBN: 9798400703348.

Von Der Leyen, Ursula (2024). *POLITICAL GUIDELINES FOR THE NEXT EUROPEAN COMMISSION 2024-2029*. Accessed 16-12-2024. URL: https://ec.europa.eu/commission/presscorner/detail/en/IP_24_926.

von der Leyen, Ursula (2024). *Mission letter to Henna Virkkunen - Executive Vice-President-designate for Tech Sovereignty, Security and Democracy*. URL: %5Curl%7Bhttps://commission.europa.eu/document/download/3b537594-9264-4249-a912-5b102b7b49a3_en?filename=Mission%5C%20letter%5C%20-%5C%20VIRKKUNEN.pdf%7D (visited on 12/16/2024).

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi (2017). "Why a right to explanation of automated decision-making does not exist in the general data protection regulation". In: *International Data Privacy Law* 7.2, pp. 76–99.

X Help Center (n.d.). *About community notes on X*. https://help.twitter.com/en/using-x/community-notes. Accessed on 12 December 2024.

YouTube Help (2024a). *Advanced search (Filters)*. URL: https://support.google.com/youtube/answer/111997?hl=en%5C&ref%5C_topic=9257501%5C&sjid=6235270588440401107-EU.

— (2024b). *Manage your recommendations and search results*. URL: https://support.google.com/youtube/answer/6342839?hl=en%5C&ref%5C_topic=9257501%5C&sjid=6235270588440401107-EU.

Zalnieriute, Monika (2021). ""Transparency Washing" in the Digital Age: A Corporate Agenda of Procedural Fetishism". In: *Critical Analysis L.* 8, p. 139.

Zhang, Xiangmin and Mark Chignell (2001). "Assessment of the effects of user characteristics on mental models of information retrieval systems". In: *Journal of the American society for information science and technology* 52.6, pp. 445–459.

Zuboff, Shoshana (2019). *The Age of Surveillance Capitalism*. Profile Books. ISBN: 9781781256855.