Towards Trustworthy Multi-stakeholder Recommender Systems

KATARINA MARCINCINOVA, Kempelen Institute of Intelligent Technologies, Slovakia
ADRIAN GAVORNIK, Kempelen Institute of Intelligent Technologies, Slovakia
MATUS MESARCIK, Kempelen Institute of Inteligent Technologies, Comenius University Bratislava, Slovakia
MICHAL KOMPAN, Kempelen Institute of Inteligent Technologies, Slovakia

Recommender system (RS) nowadays has to reflect, combine, and solve often contrasting requirements and expectations from dozens of stakeholders. In fact, the problem is even more complicated as each of the stakeholders can simultaneously have various objectives creating a diverse and complicated environment with several stakeholders and their objectives in which a RS operates. To address this problems the Multi-objective and Multi-stakeholder RS have been studied in the literature. However, a limited attention has been paid to trustworthiness aspects of such RS, which slowly becomes a new standard. In this paper, we highlight open challenges and important questions which need to be addressed on the way to trustworthy Multi-stakeholder RS.

Keywords: Recommender systems, Trustworthy RS, Multi-stakeholder, Multi-objective

Reference Format:
Katarina Marcincinova, Adrian Gavornik, Matus Mesarcik, and Michal Kompan. 2025. Towards Trustworthy Multi-stakeholder Recommender Systems. In *Proceedings of Fourth European Workshop on Algorithmic Fairness (EWAF'25).* Proceedings of Machine Learning Research, 6 pages.

1 Introduction and Related Works

Recommender systems (RSs) have tremendously impacted the online life of every user on the Web. Starting with the Goldberg's mail recommendation system Tapestry in 1992, which aimed at routing relevant mail to competent employees [13], the field evolved to serving billions of requests for recommendations every day on social media or e-commerce platforms. The role of a regular user on the Web also evolved. The user generated content paradigm transformed a strict consumers to the role of "prosumers", which mixes (or switches between) two roles of a consumer and producers. Last but not least, large platforms, e.g., YouTube, Facebook, Amazon empower users to access content from many producers (e.g., businesses). This was reflected by introducing the Multi-Stakeholder (MSRS) or Multi-Objective (MORS) recommender system. While the latter does not necessarily involve different stakeholders (a user may have several objectives simultaneously), e.g., to spend a free time on social network while to educate about new concepts, this is usually not the case with multiple stakeholders (different stakeholders tend to have different, yet still multiple objectives). Despite the MORSs and MSRSs have been studied in the literature,

Authors' Contact Information: Katarina Marcincinova, Kempelen Institute of Intelligent Technologies, Slovakia, name.surname@kinit.sk; Adrian Gavornik, Kempelen Institute of Intelligent Technologies, Slovakia, name.surname@kinit.sk; Matus Mesarcik, Kempelen Institute of Inteligent Technologies, Comenius University Bratislava, Slovakia, name.surname@kinit.sk; Michal Kompan, Kempelen Institute of Inteligent Technologies, Slovakia, name.surname@kinit.sk.

EWAF'25, June 30–July 02, 2025, Eindhoven, NL

a limited attention has been paid to trustworthiness of such RS. In this paper, we analyze and emphasize open challenges and important questions which need to be addressed on the way to trustworthy MSRSs.

Although RSs are typically framed as user-centric tools designed to enhance individual experiences through personalized content, these systems are increasingly recognized as complex socio-technical assemblages that affect multiple stakeholders at once [11, 20, 22, 27]. Some [3] argue that the widespread use of RSs can have a serious impact on how humans receive information and ultimately perceive the world. Not only that, but the increasing reliance on RSs can affect our moral skills [24] and even reshape human agency as such [14]. This is important, especially in the context of very large social media platforms, where users receive information and perceive content on a daily basis, which has often been abused (e.g., for political campaigns) [19]. Therefore, this socio-technical understanding of RSs goes beyond the mere technological artifacts and instead acknowledges and makes explicit its entanglement with the social, cultural and political dimensions of its development and deployment. For this reason, the question of how to create ethical or trustworthy RS is becoming even more important.

In the literature, there is no single accepted definition of "trustworthiness" with respect to AI systems, but rather we find specific requirements that apply to it. For example, according to EGTAI [4], "trustworthy" AI should be lawful, ethical, and robust. Specifically, seven key requirements for trustworthy AI systems have been formulated: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, nondiscrimination and fairness; societal and environmental being; and accountability.

The discourse on trustworthy RS has mainly focused on the negative aspect or definition, i.e., by identifying and minimizing risks, such as biases, unfairness, or privacy concerns, including other harms through specific metrics and technological fixes [1, 20]. The state-of-the-art MORSs and MSRSs in the context of trustworthy also usually focus on individual requirements of trustworthiness mentioned above. For example, Schedl et al. [22] identified three main dimensions of trustworthy RS - bias, fairness and non-discrimination; transparency and explainability; privacy and security. Others have focused on specific requirement, such as diversity [16, 25] or fairness [9, 21].

It could be argued that current approaches omit the positive aspect or the definition of trustworthiness, one of which is proactive advancement of human and especially ethical values in the design of RSs as proposed by Stray [27]. More importantly, in the discussion about MSRSs, trustworthiness is not addressed, as there is a lack of consensus and research exploring the key implications of such systems [8].

We aim to address this gap by expanding the concept of trustworthy MSRSs to incorporate both negative (risk mitigation) and positive (value-sensitive) aspects. We also argue that following the tripartite structure of trustworthy AI (legal compliance, robustness, adherence to ethical norms) should be adopted to how RSs can be in alignment with fundamental rights and societal values, particularly in the European context where regulatory frameworks like the Digital Services Act (DSA) [7], consumer-related legislation such as Unfair Commercial Practices Directive (UCPD) [5], and Consumer Rights Directive (CRD) [6] provide normative guidance. We also argue that trustworthiness comes not only from technical solutions, but from the balanced consideration of various stakeholder values, objectives, and the inevitable tensions that arise between them.

2 Stakeholders and Objectives

The wide variety of stakeholders are involved and affected by the RSs - from individual users to societies as a whole. The task of identifying and conceptualizing stakeholders is therefore not trivial. As Milano et al. [20] discuss, this concerns the mismatch in ontologies between the system design and the actual stakeholders it has an impact on.

They argue that this makes ethical evaluations of RSs rather challenging. This task becomes even more complex in the context of legal frameworks. There are notable overlaps between how various regulatory regimes define and categorize stakeholders. Additionally, the appropriate level of abstraction for defining stakeholders varies across contexts - from individual users, user communities, to broader societal interests. We focus primarily on RSs for social media platforms, but other areas, such as e-commerce or streaming services, are sharing similar properties.

In the area of very large social media platforms (i.e., Instagram, Facebook, or TikTok), the most prominent stakeholders are [11, 22, 27]: end users (consumers), content producers (creators), and service providers (social media platforms). However, there are other stakeholders that are equally affected by RSs, including developers, regulators or policymakers, and even society as a whole. The ongoing discourse in the context of trustworthy RSs explores whether the items should be included as a separate stakeholder [30]. In other words, has a recommended item a “right” to be recommended? We argue that it should not be considered as a separate stakeholder. However, the item’s “rights” come naturally from the trustworthy definition and properties (i.e., diversity, novelty, or fairness).

It is obvious that the objectives of the stakeholders differ. For example, as a major stakeholder in RSs, **users (consumers)** are primarily motivated to fulfill their information needs [23]. These can be further translated into various objectives, such as finding new friends, having fun in the case of social media platforms, or buying a product with the best price and required quality in the case of e-commerce [15]. Moreover, users benefit from regulatory requirements for providers of (very large) platforms, e.g., under the DSA [7] that should also be considered as part of their objectives. These relate to transparency regarding the ranking parameters of RSs, or users should be able to modify the parameters and explanations of the main parameters. These shall at least include the most significant criteria for recommendations and reasons for the relative importance of those parameters [7, Article 27].

On the other hand, it is clear that different **content producers** as another stakeholder follow their own objectives, as maximizing the amount of content consumed or increasing the number of subscribed end-users on a social media platforms; or selling out outdated products [18] or increasing customer loyalty in the case of e-commerce.

Finally, as a third core stakeholder group, **providers** (i.e., social media platforms), are the most powerful element as they possess all the information and can directly impact other groups. Clearly, providers represent a key element that has to ensure the alignment to the trustworthy aspects of RSs. Among the typical objectives of providers are: maximizing the profit, increasing the dwell-time, or reducing the churn-rate. In other non-commercial applications, these can be, i.e., increasing knowledge, improving health, or recommending most enjoyable route in a museum [2]. Platforms face the most comprehensive regulatory obligations, with the DSA requiring disclosure of RS parameters, user modification options, and non-profiling alternatives [7, Article 38] for very large platforms.

However, from the perspective of trustworthy MSRSs, it is necessary to consider the objectives of other stakeholders affected by the system. We should consider the developers of these systems, whose objectives are mostly connected to the technical properties of the RSs, e.g., recommendation complexity or computation time [17]. Similarly, regulators and policymakers aim to set the requirements for other stakeholder groups to follow. For instance, in the case of social media platforms, an expandability of main parameters of RS [10] can be required. Following the regulations and/or trustworthiness of RSs, it is necessary to assess whether the RS meets the required criteria. Thus, auditors, as another stakeholder affected by RSs, focus on the degree of auditability of specific RS [26]. Last but not least, we might consider the objectives of society as a whole, which might include general well-being, democratic values, and rule of law. However, it is important to note that the objectives of society will vary, especially when we include cultural considerations in different parts of the world.

Given that every stakeholder has its own objectives in relation to RSs, various value tensions can arise between them [22, 27]. In other words, there are trade-offs not only between different stakeholders and their objectives, but also between the different values. For example, content creators may want to promote specific values on social media platforms, such as a healthy lifestyle, i.e., through dietary videos or sponsored content. Such content may not seem problematic, but when we start to promote it to individuals, such as those who suffer from eating disorder, the stream of such videos may have a negative impact on their well-being [27]. Another value tension can be found between the platforms obligation stemming (Article 27 of DSA) to "disclose the main parameters used in their recommender systems" and the platform objective to protect such attributes of the RSs because of a trade secret.

Several authors [1, 9, 11, 20, 28] have pointed that through MSRSs and MORSs, we may find balance between objectives and values of multiple stakeholders and at the same time enhance our understanding of the broader societal impacts, inc. cultural considerations. We argue that research on trustworthy MSRSs is needed, especially when discussing the positive aspect of trustworthiness, which seeks to emphasize the importance of human values.

3 Future Directions

Trustworthy MORS. The computer science literature has been primarily concerned with the question what MSRSs and MORSs are and how to create such systems for particular applications [29]. However, it has not delved deeper into specific solutions, especially when considering not only the complexity of the stakeholders involved and influenced by them, but also the trustworthy aspects. For instance, the fairness, widely researched in standard recommenders, needs a special attention in the MORS setting (e.g., which stakeholders should be treated fair).

Human values. Another challenge is how to operationalize human values in the design of MSRSs. It is difficult to measure human values as a concrete definition is missing (i.e., "well-being" is contextual and varies among different stakeholders). There are prerequisites (i.e., techniques to measure adherence to values) that are important to consider, but specific methods and solutions are needed to incorporate values of stakeholders at once [27]. We suggest that case-based and context-sensitive ethical assessments can move beyond abstract ethical principles by including systematic stakeholder identification [12]. However, these methods must go beyond identifying stakeholders - they should capture the interplay of stakeholder values, objectives, and potential tension in between.

Computational complexity. It is clear, that the multiple objectives coming from different stakeholders can bring overhead from the computation complexity point of view. Especially when a conflicting objectives are on the table, some kind of guidance is required. Incorporating human values can on the one side help to deal with such conflicting objectives. On the contrary, additional objectives are introduced by the trustworthy requirements, increasing the computational complexity of such RSs, demeaning novel technical solutions.

Regulatory requirements. The complex regulatory environment surrounding RSs complicates the development of trustworthy MRSRs. It requires careful consideration of the legal obligations while balancing technical capabilities and stakeholder interests. For instance the DSA requires the platform to communicate reasons for the relative importance of main parameters used for recommendation, which may be in case of MORS and MSRSs challenging.

Acknowledgments

Partially funded by the NextGenerationEU through the Recovery and Resilience Plan for Slovakia, project No. 09I03-03-V04-00336; and by European Union, project Low Resource Artificial Intelligence, GA No. 101136646.

References

[1] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. 2020. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction* 30, 1 (March 2020), 127–158. doi:10.1007/s11257-019-09256-1

[2] Alba Amato. 2023. Recommender Systems in the Museum Sector: An Overview. In *Advanced Information Networking and Applications*, Leonard Barolli (Ed.). Springer International Publishing, Cham, 251–260.

[3] Maximilian Boeker and Aleksandra Urman. 2022. An Empirical Investigation of Personalization Factors on TikTok. In *Proceedings of the ACM Web Conference 2022 (WWW '22)*. ACM, New York, NY, USA, 2298–2309. doi:10.1145/3485447.3512102

[4] European Commission, Content Directorate-General for Communications Networks, and Technology. 2019. *Ethics guidelines for trustworthy AI*. Publications Office, Brussels, Belgium. doi:10.2759/346720

[5] Council of European Union. 2022. Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market (Unfair Commercial Practices Directive) (Consolidated version: 28 May 2022). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02005L0029-20220528 Official Journal of the European Union L 149, 11.6.2005, p. 22–39.

[6] Council of European Union. 2022. Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights (Consumer Rights Directive) (Consolidated version: 28 May 2022). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02011L0083-20220528 Official Journal of the European Union L 304, 22.11.2011, p. 64–88.

[7] Council of European Union. 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). https://eur-lex.europa.eu/eli/reg/2022/2065/oj Official Journal of the European Union L 277, 27.10.2022, p. 1–102.

[8] Yashar Deldjoo, Alejandro Bellogin, and Tommaso Di Noia. 2021. Explaining recommender systems fairness and accuracy through the lens of data characteristics. *Information Processing & Management* 58, 5 (2021), 102662. doi:10.1016/j.ipm.2021.102662

[9] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, and Dario Zanzonelli. 2024. Fairness in recommender systems: research landscape and future directions. *User Modeling and User-Adapted Interaction* 34, 1 (March 2024), 59–108. doi:10.1007/s11257-023-09364-z

[10] Matteo Fabbri. 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?. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) *(AIES '23)*. Association for Computing Machinery, New York, NY, USA, 1002–1003. doi:10.1145/3600211.3604749

[11] Matteo Fabbri. 2023. Social influence for societal interest: a pro-ethical framework for improving human decision making through multi-stakeholder recommender systems. *AI & SOCIETY* 38, 2 (April 2023), 995–1002. doi:10.1007/s00146-022-01467-2

[12] Adrian Gavornik, Juraj Podroužek, Matus Mesarcik, Sara Solarova, Stefan Oresko, and Maria Bielikova. 2022. Utilising the Assessment List for Trustworthy AI: Three Areas of Improvement. In *2nd Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies - AIofAI 2022*, Vol. 3275. CEUR-WS.org, Vienna, Austria, 8. https://ceur-ws.org/Vol-3275/paper3.pdf

[13] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. 1992. Using collaborative filtering to weave an information tapestry. *Commun. ACM* 35, 12 (Dec. 1992), 61–70. doi:10.1145/138859.138867

[14] Ville Jylhä, Noora Hirvonen, and Jutta Haider. 2024. Algorithmic recommendations in the everyday life of young people: imaginaries of agency and resources. *Information, Communication & Society* 0, 0 (2024), 1–17. doi:10.1080/1369118X.2025.2470227

[15] Michal Kompan, Peter Gaspar, Jakub Macina, Matus Cimerman, and Maria Bielikova. 2022. Exploring Customer Price Preference and Product Profit Role in Recommender Systems. *IEEE Intelligent Systems* 37, 1 (2022), 89–98. doi:10.1109/MIS.2021.3092768

[16] Alexandre Letard, Nicolas Gutowski, Olivier Camp, and Tassadit Amghar. 2024. Bandit algorithms: A comprehensive review and their dynamic selection from a portfolio for multicriteria top-k recommendation. *Expert Systems with Applications* 246 (2024), 123151. doi:10.1016/j.eswa.2024.123151

[17] Shiwei Li, Huifeng Guo, Xing Tang, Ruiming Tang, Lu Hou, Ruixuan Li, and Rui Zhang. 2024. Embedding Compression in Recommender Systems: A Survey. *ACM Comput. Surv.* 56, 5, Article 130 (Jan. 2024), 21 pages. doi:10.1145/3637841

[18] Sichun Luo, Chen Ma, Yuanzhang Xiao, and Linqi Song. 2023. Improving Long-Tail Item Recommendation with Graph Augmentation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (Birmingham, United Kingdom)

(CIKM '23). Association for Computing Machinery, New York, NY, USA, 1707–1716. doi:10.1145/3583780.3614929

[19] Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok. In *Proceedings of the 12th ACM Conference on Web Science* (Southampton, United Kingdom) *(WebSci '20)*. Association for Computing Machinery, New York, NY, USA, 257–266. doi:10.1145/3394231.3397916

[20] Silvia Milano, Mariarosaria Taddeo, and Luciano Floridi. 2021. Ethical aspects of multi-stakeholder recommendation systems. *The Information Society* 37, 1 (Jan. 2021), 35–45. doi:10.1080/01972243.2020.1832636

[21] Naime Ranjbar Kermany, Weiliang Zhao, Jian Yang, Jia Wu, and Luiz Pizzato. 2021. A fairness-aware multi-stakeholder recommender system. *World Wide Web* 24, 6 (Nov. 2021), 1995–2018.

[22] Markus Schedl, Vito Walter Anelli, and Elisabeth Lex. 2025. *Technical and Regulatory Perspectives on Information Retrieval and Recommender Systems: Fairness, Transparency, and Privacy*. The Information Retrieval Series, Vol. 50. Springer Nature Switzerland, Cham. doi:10.1007/978-3-031-69978-8

[23] Fabian Schneider, Anja Feldmann, Balachander Krishnamurthy, and Walter Willinger. 2009. Understanding online social network usage from a network perspective. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement* (Chicago, Illinois, USA) *(IMC '09)*. Association for Computing Machinery, New York, NY, USA, 35–48. doi:10.1145/1644893.1644899

[24] Nick Schuster and Seth Lazar. 2025. Attention, moral skill, and algorithmic recommendation. *Philosophical Studies* 182, 1 (Jan. 2025), 159–184. doi:10.1007/s11098-023-02083-6

[25] Rahul Shrivastava, Dilip Singh Sisodia, and Naresh Kumar Nagwani. 2023. On Diverse and Serendipitous Item Recommendation: A Reinforced Similarity and Multi-objective Optimization-Based Composite Recommendation Framework. In *Machine Intelligence Techniques for Data Analysis and Signal Processing*, Dilip Singh Sisodia, Lalit Garg, Ram Bilas Pachori, and M. Tanveer (Eds.). Springer Nature Singapore, Singapore, 1–13.

[26] Ivan Srba, Robert Moro, Matus Tomlein, Branislav Pecher, Jakub Simko, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, Adrian Gavornik, and Maria Bielikova. 2023. Auditing YouTube's Recommendation Algorithm for Misinformation Filter Bubbles. *ACM Trans. Recomm. Syst.* 1, 1, Article 6 (Jan. 2023), 33 pages. doi:10.1145/3568392

[27] Jonathan Stray, Alon Halevy, Parisa Assar, Dylan Hadfield-Menell, Craig Boutilier, Amar Ashar, Chloe Bakalar, Lex Beattie, Michael Ekstrand, Claire Leibowicz, Connie Moon Sehat, Sara Johansen, Lianne Kerlin, David Vickrey, Spandana Singh, Sanne Vrijenhoek, Amy Zhang, McKane Andrus, Natali Helberger, Polina Proutskova, Tanushree Mitra, and Nina Vasan. 2024. Building Human Values into Recommender Systems: An Interdisciplinary Synthesis. *ACM Trans. Recomm. Syst.* 2, 3 (2024), 20:1–20:57. doi:10.1145/3632297

[28] Jonathan Stray, Ivan Vendrov, Jeremy Nixon, Steven Adler, and Dylan Hadfield-Menell. 2021. What are you optimizing for? Aligning Recommender Systems with Human Values. arXiv:2107.10939 [cs.IR] https://arxiv.org/abs/2107.10939

[29] Fatima Ezzahra Zaizi, Sara Qassimi, and Said Rakrak. 2023. Multi-objective optimization with recommender systems: A systematic review. *Information Systems* 117 (2023), 102233. doi:10.1016/j.is.2023.102233

[30] Yong Zheng and Juan Ruiz Toribio. 2021. The role of transparency in multi-stakeholder educational recommendations. *User Modeling and User-Adapted Interaction* 31, 3 (01 Jul 2021), 513–540. doi:10.1007/s11257-021-09291-x