

# Report on NORMalize: The Second Workshop on the Normative Design and Evaluation of Recommender Systems

Alain Starke<sup>1,2</sup>, Sanne Vrijenhoek<sup>1</sup>, Lien Michiels<sup>3,6</sup>, Johannes Kruse<sup>4,7</sup> and Nava Tintarev<sup>5</sup>

<sup>1</sup>University of Amsterdam, Amsterdam, the Netherlands

<sup>2</sup>University of Bergen, Bergen, Norway

<sup>3</sup>University of Antwerp, Antwerp, Belgium

<sup>4</sup>JP/Politikens Media Group, Copenhagen, Denmark

<sup>5</sup>Maastricht University, Maastricht, the Netherlands

<sup>6</sup>imec-SMIT, Vrije Universiteit Brussel, Brussels, Belgium

<sup>7</sup>Technical University of Denmark, Kongens Lyngby, Denmark

## Abstract

Recommender systems are among the most widely used applications of artificial intelligence. Because of their widespread use, it is important that practitioners and researchers think about the impact they may have on users, society, and other stakeholders. To that effect, the NORMalize workshop seeks to introduce *normative thinking*, to consider the norms and values that underpin recommender systems in the recommender systems community. The objective of NORMalize is to bring together a growing community of researchers and practitioners across disciplines who want to think about the norms and values that should be considered in the design and evaluation of recommender systems, and further educate them on how to reflect on, prioritise, and operationalise such norms and values. This document is a report on the second NORMalize workshop, co-located with ACM RecSys '24 in Bari, Italy.

## Keywords

normative thinking, normative design, recommender systems, norms, values, value-sensitive design

## 1. Introduction

The possible societal impact of recommender systems is becoming increasingly important for the systems' designers [1]. This is underlined by the increased importance of so-called 'beyond-accuracy' metrics in recommender research. These include methods that devote attention to notions of fairness, such as statistical parity or equality of opportunity in the design and evaluation of recommender systems [2, 3]. However, this also means that many values could be

---

*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*

✉ a.d.starke@uva.nl (A. Starke); s.vrijenhoek@uva.nl (S. Vrijenhoek); lien.michiels@uantwerpen.be (L. Michiels); johannes.kruse@eb.dk (J. Kruse); n.tintarev@maastrichtuniversity.nl (N. Tintarev)

🆔 0000-0002-9873-8016 (A. Starke); 0000-0002-1031-4746 (S. Vrijenhoek); 0000-0003-0152-2460 (L. Michiels); 0009-0007-5830-0611 (J. Kruse); 0000-0003-1663-1627 (N. Tintarev)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

considered when developing recommender systems, of which fairness towards the end-users of the system is only but one example [4].

Identifying and balancing values of recommender systems requires so-called *normative thinking* and decision-making [5, 6, 7]. Normative thinking requires recommender designers to reflect on how or what the system *should be*, rather than focusing on what the current state of the system (output) *is*. Beyond identifying relevant values, this also includes determining how these values would be present in what is recommended by a system, examining possible conflicts between different values, and justifying how certain values in specific cases should be prioritised over others [8].

Last year saw the first edition of our workshop. We organized an interactive session in which attendees were encouraged to come up with their own normative framework for a specific use case. Besides that, we also welcomed our first research contribution and proceedings [9], publishing nine research papers.

## 2. Overview of Contributions

This year's workshop continued last year's work by again welcoming original research contributions. These are included in the workshop's proceedings, describing new research on the design and evaluation of normative recommenders. In total, we received 9 paper submission, of which 6 were accepted for the proceedings.

The NORMalize2024 program consisted of two blocks with three research presentations each, and a few interactive parts inbetween. The first block was a session on 'Data and Framework', featuring the following presentations:

- IDEA - Informfully Dataset with Enhanced Attributes  
*Lucien Heitz, Nicolas Mattis, Oana Inel, and Wouter van Atteveldt*
- From Walls to Windows: Creating Transparency to Understand Filter Bubbles in Social Media  
*Luka Bekavac, Kimberly Garcia, Jannis Strecker, Simon Mayer, and Aurelia Tamo-Larrieux*
- Generating Diverse Synthetic Data Sets for Evaluation of Real-life Recommender Systems  
*Miha Malenšek, Blaž Škrlić, Blaž Mramor, and Jure Demšar*

Each block was also followed by a group discussion. This allowed us to synthesize insights and to foster discussion between attendees. The second block was a session on 'Policy and Values', featuring the following presentations:

- Diversifying for Democracy: Cultivating Publics via Algorithmic Design and the Normative Consequences for Journalism  
*Jannie Møller Hartley and Elisabetta Petrucci*
- Navigating the Digital Services Act: Scenarios of Transparency and Control in VLOP Recommender Systems  
*Urbano Reviglio and Matteo Fabbri*
- Value Identification in Multi-Stakeholder Recommender Systems for Humanities and Historical Research: The Case of the Digital Archive Monasterium.net  
*Florian Atzenhofer-Baumgartner, Bernhard Geiger, Georg Vogeler, and Dominik Kowald*

### 3. Disagreeenotes

This year's program featured short, provocative statements by members of the organizing committee. We named these *Disagreeenotes*, as we expected a part of the attendees to disagree with our viewpoints, even though they might be held by some members of the RecSys community.

The goal was to foster discussion among the attendees on propositions relevant to the workshop. The workshop organizers ensured that they could convincingly argue both positions on the statements, to facilitate a discussion in case the audience all shared the same viewpoints. This had as an added benefit that it created a safe space, where perspectives were not taken personally. All four disagreeenotes sparked lively discussion among the participants of the workshop. We wish to thank the participants for their active participation in these insightful discussions. Below, we summarize the presented disagreeenotes, as well as the main insights raised in the subsequent discussions.

**Disagreeenote 1: We do not need personalized recommender systems.** The first Disagreeenote triggered a lot of interaction from the audience, and led to a discussion that almost at the philosophical level dissected what it means to be 'personalized' and what it means to 'need' something. For example, it was noted that we do not need personalized recommender systems in the same sense that we need water and food. While personalization can be considered helpful to filter through large amounts of information, other non-personalized alternatives may be possible, and sometimes even preferred. For example, in the context of news, it is important that some parts of the online news platform is and remains curated by editors, as it is important that some news reaches everyone. Yet, at the same time, personalization can be very beneficial to surface news that may otherwise never make it onto the homepage, such as regional news. To summarize, when building recommender systems, we should evaluate what needs or desires they address, and whether these needs and desires may not be better served by a non-personalized system.

**Disagreeenote 2: There is no such thing as unbiased data, therefore, striving for unbiased AI is nonsense** While the audience agreed that data is inherently biased and that striving for unbiased data is an unrealistic goal, the second part of the statement prompted discussion. Data collected from the real world reflects human biases, prompting the question of what objectives should guide the development of AI systems. Should the focus be on achieving "unbiased" AI, or is it more pragmatic to prioritize transparency and effective bias mitigation? Transparency regarding how data is collected, whom it represents, and the context of its use can enable practitioners to better interpret and responsibly leverage data, even when it is biased. The discussion also examined the societal risks of ignoring bias, such as reinforcing systemic inequalities, and considered the allocation of responsibility: should developers bear the primary responsibility, or should users and other stakeholders share this burden? This Disagreeenote underscored the inherent complexity of striving for fairness and accountability in AI.

**Disagreeenote 3: Ethical Guidelines, and non-binding types of policy, are as far as government bodies should go to regulate recommender systems** If one of the key points

of NORMalize is to discover latent norms and values that we are often not even consciously aware of, then so must we recognize that European laws such as the AI Act and the Digital Service Act embody European norms and values, that are now imposed on the rest of the world. While during the main conference there was often a good deal of muttering about these laws, and specifically GDPR, participants of the NORMalize workshop were (perhaps unsurprisingly) generally in favor of increased regulation. They noted that there is no evidence yet that regulation hinders innovation, but also that laws need to be well-structured and clear in order to be effective.

**DisagreeNote 4: There are too many workshops about roughly the same topic. NORMalize should not be organized next year to allow other workshops to gain more critical mass.** This disagreeNote was meant to entice participants to share thoughts about potential future directions NORMalize could take. RecSys’24 hosted 21 workshops. Out of those, FAccTRec, AltRecSys and RecSoGood were topically strongly related to NORMalize, whereas domain-specific workshops such as INRA, MuRS or HealthRec could have benefited from participants taking a normative perspective. As workshop organizers, we wondered whether we, as one of the smaller workshops, should take a step back, and allow other workshops to gain critical mass, and effectuate change in the conference at large. Participants saw the merit of the point, yet also argued that NORMalize was quite original in its setup, and likely the only workshop that succeeded in bringing interdisciplinary perspectives to the conference.

## 4. Submitted Work

The accepted work (9 registered abstracts, 6 accepted) can be thematically clustered into papers dealing with “Data and Frameworks” and “Policy and Values”. Each paper received three reviews by members of the program committee, at least one of which was from a technical- and one from a social science/humanities background.

### 4.1. Data and Frameworks

Publicly available datasets are crucial for addressing challenges in recommender systems, particularly concerning content diversity and user behavior analysis. In their work, “IDEA - Informfully Dataset with Enhanced Attributes”, Heitz et al. introduced the IDEA dataset—an open-source collection that combines diverse news articles, detailed user profiles, item recommendations, and rich user-item interactions from a field study on news consumption. This dataset integrates real-time session tracking with self-reported survey data on user satisfaction and knowledge acquisition, providing a valuable resource for designing normative recommender systems.

Continuing the theme of content diversity, Bekavac et al. presented “From Walls to Windows: Creating Transparency to Understand Filter Bubbles in Social Media”. They developed SOAP (System for Observing and Analyzing Posts), a novel system that leverages a multimodal language model to study filter bubbles at scale on large online platforms. SOAP can generate and navigate filter bubbles based on topic prompts, enabling analysis of how topic diversity

diminishes over time in social media feeds. Their findings reveal a significant decline in topic diversity within just 60 minutes of scrolling, highlighting the impact of recommender systems on content diversity.

Further contributing to resources for recommender system evaluation, Malenšek et al. introduced “Generating Diverse Synthetic Datasets for Evaluation of Real-life Recommender Systems”. They developed a framework for generating synthetic datasets that are diverse and statistically coherent, tailored to real-world recommender systems. This approach allows for controlled creation of datasets with customizable attributes, such as complex feature interactions and specific distributions, facilitating experiments that require specific experimental setups. Their modular and open-source Python package addresses the need for flexible synthetic data generation, aiding in benchmarking algorithms, detecting bias, and advancing recommender system evaluations.

## 4.2. Policy and Values

Policy surrounding recommender systems and their values can take many forms. On the one hand, legislation can help to safeguard against the introduction of harmful norms and values and to set standards. On the other hand, designers and practitioners of relevant systems can define which norms and values should be incorporated into their platforms.

One example is found in journalism. Møller Hartley and Petrucci show in their work titled *Diversifying for Democracy: Cultivating Publics via Algorithmic Design and the Normative Consequences for Journalism* how the concept of diversity, which is an often used value in news recommender systems [4], is typically rooted in two related concepts: filter bubbles and choice overload. Their literature review suggests that solutions to diversity problems can therefore be sought in exposure and viewpoint diversity. One example provided in the paper is that recommending ‘more of the same’ could not only be boring to users, but also dangerous to democratic processes.

A different perspective is given by law researchers. Reviglio and Fabbri examine how EU law could affect large platforms, in their work *Navigating the Digital Services Act: Scenarios of Transparency and Control in VLOP Recommender Systems*. Their work discusses how the Digital Services Act affects various platforms that run recommender system services, particularly those on large platforms. It highlights which parts of the EU legislation contain normative grounds and what the minimal and maximum conditions are for different forms of personalization and the collection of personal data.

Finally, the work of Atzenhofer-Baumgartner et al. showcases an example of value identification in a digital archive. Their work titled *Value Identification in Multi-Stakeholder Recommender Systems for Humanities and Historical Research: The Case of the Digital Archive Monasterium.net* shows how various stakeholders and users of this digital archive differ in their main values. For example, editors of this platform value visibility of different content, while researchers would like recommendations to be relevant to them, focusing on accuracy. The work discusses main challenges, for example with regard to conflicting values.

### 4.3. Conclusion

These two blocks show the versatility of the topics concerning normativity and recommender systems. We feel that the scope of this topic is not limited to the contributions we received this year, but that it does provide insights on how norms and values are related to recommender system design. We wholeheartedly invite you to read these proceedings and, if possible, to contribute to a future edition of this workshop.

### Acknowledgments

We would like to thank the participants and authors of accepted contributions for their valuable inputs to the workshop, our program committee for their thoughtful reviews, as well as the RecSys'24 organisers for their support in the organisation of NORMalize. Finally, we would like to thank our employers and funding bodies. Sanne Vrijenhoek's contribution to this research is supported by the AI, Media and Democracy Lab. Lien Michiels' contribution to this research was supported by the Research Foundation Flanders (FWO) under grant number S006323N and the Flanders AI research program. Johannes Kruse's contribution to this research is supported by the Innovation Foundation Denmark under grant number 1044-00058B and Platform Intelligence in News under project number 0175-00014B. Alain Starke's contribution was in part supported by the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the Centre for Research-based Innovation scheme, project number 309339. Nava Tintarev's contribution is supported by the project ROBUST: Trustworthy AI-based Systems for Sustainable Growth with project number KICH3.LTP.20.006, which is (partly) financed by the Dutch Research Council (NWO), RTL, and the Dutch Ministry of Economic Affairs and Climate Policy (EZK) under the program LTP KIC 2020-2023. *All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.*

### References

- [1] M. D. Ekstrand, M. Tian, M. R. I. Kazi, H. Mehrpouyan, D. Kluver, Exploring author gender in book rating and recommendation, in: Proceedings of the 12th ACM conference on recommender systems, 2018, pp. 242–250.
- [2] R. Mehrotra, J. McInerney, H. Bouchard, M. Lalmas, F. Diaz, Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems, in: Proceedings of the 27th acm international conference on information and knowledge management, 2018, pp. 2243–2251.
- [3] E. Purificato, L. Boratto, E. W. De Luca, Do graph neural networks build fair user models? assessing disparate impact and mistreatment in behavioural user profiling, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 4399–4403.
- [4] S. Vrijenhoek, G. Bénédicte, M. Gutierrez Granada, D. Odijk, M. De Rijke, Radio – rank-aware divergence metrics to measure normative diversity in news recommendations, in:

Proceedings of the 16th ACM Conference on Recommender Systems, RecSys '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 208–219. URL: <https://doi.org/10.1145/3523227.3546780>. doi:10.1145/3523227.3546780.

- [5] S. Buckler, Normative theory, *Theory and methods in political science* 3 (2010) 156–180.
- [6] J. J. Thomson, *Normativity*, 2010.
- [7] T. A. Christiani, Normative and empirical research methods: Their usefulness and relevance in the study of law as an object, *Procedia-Social and Behavioral Sciences* 219 (2016) 201–207.
- [8] B. C. Stahl, Morality, ethics, and reflection: a categorization of normative is research, *Journal of the association for information systems* 13 (2012) 1.
- [9] S. Vrijenhoek, L. Michiels, J. Kruse, A. Starke, J. V. Guerrero, N. Tintarev, Report on normalize: The first workshop on the normative design and evaluation of recommender systems, in: *Proceedings of the First Workshop on the Normative Design and Evaluation of Recommender Systems (NORMalize 2023)*, co-located with the 17th ACM Conference on Recommender Systems (RecSys 2023), volume 3639, CEUR, 2023.