# Second Workshop: Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2022)

Eva Zangerle
eva.zangerle@uibk.ac.at
Universität Innsbruck
Austria

Christine Bauer
c.bauer@uu.nl
Utrecht University
The Netherlands

Alan Said
University of Gothenburg
Sweden
alansaid@acm.com

## ABSTRACT

Evaluation of recommender systems is a central activity when developing recommender systems, both in industry and academia. The second edition of the PERSPECTIVES workshop held at RecSys 2022 brought together academia and industry to critically reflect on the evaluation of recommender systems. In the 2022 edition of PERSPECTIVES, we discussed problems and lessons learned, encouraged the exchange of the various perspectives on evaluation, and aimed to move the discourse forward within the community. We deliberately solicited papers reporting a reflection on problems regarding recommender systems evaluation and lessons learned. The workshop featured interactive parts with discussions in small groups as well as in the plenum, both on-site and online, and an industry keynote.

## CCS CONCEPTS

• **Information systems → Personalization**; **Recommender systems**; **Evaluation of retrieval results**; • **Human-centered computing → HCI design and evaluation methods**.

## KEYWORDS

evaluation, personalized systems, recommender systems, methods, information retrieval

**ACM Reference Format:**
Eva Zangerle, Christine Bauer, and Alan Said. 2022. Second Workshop: Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2022). In *Sixteenth ACM Conference on Recommender Systems (RecSys '22), September 18–23, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3523227.3547408

## 1 INTRODUCTION

Evaluation demands attention throughout every phase in a recommender system's lifecycle—in design and development as well as for continuous improvement while in operation. The evaluation of a recommender system may focus on an algorithmic, system-centric perspective (e.g., the predictive accuracy of recommendation algorithms), the user's perspective in a user-centric evaluation (e.g., how users assess its quality or experience when interacting with a recommendation system), or both perspectives [6, 7]. That said,

the evaluation of recommender systems covers a wide spectrum of different aspects, both in academia and in industry. This observation was the motivation and starting point for the PERSPECTIVES workshop series.

The goal of the PERSPECTIVES workshop series is to capture the current state of evaluation and determine whether there is or should be a different target that recommender system evaluation should strive for. Building on the first edition of the PERSPECTIVES workshop, this second edition of the workshop continued the discussion and particularly addressed the question of where we should go as a community.

From the discussion in the first PERSPECTIVES workshop held at RecSys 2021, it became clear that despite all the ongoing efforts in the field of evaluating recommender systems, there are still many open issues that we need to tackle in the recommender systems research and practitioner community [10]. In the second edition of the PERSPECTIVES workshop held at RecSys 2022, we continued the discussions to bring consensus and consolidation into an ever-increasingly active field of recommender systems research. The discussions in the second PERSPECTIVES workshop substantiated that we need to take and embrace a wide(r) scope of perspectives, covering the full spectrum of factors relevant when assessing the quality of recommender systems: from evaluation methods and experimental designs [3], impact and purposes [4, 5], levels of maturity of the system employed, the context of the recommender system (industry or academia), to stakeholders affected [1, 2] and fairness considerations [8], just to name a few of those central factors.

## 2 TOPICS OF INTEREST AND MATERIAL

The workshop solicited papers addressing topics such as those listed below.

- Case studies of difficult, hard-to-evaluate scenarios
- Evaluations with contradicting results
- Showcasing (structural) problems in recommender systems evaluation
- Integration of offline and online experiments
- Multi-Stakeholder evaluation
- Divergence between evaluation goals and what is actually captured by the evaluation
- Nontrivial and unexpected experiences from practitioners

Aside from these core topics and the papers submitted by participants, we sought to gather feedback from the participants before and during the workshop regarding pressing issues related to the evaluation of recommender systems that should be addressed in the workshop. Hence, the topics discussed during the workshop extended beyond the list of topics above. In addition, we deliberately

solicited papers reporting problems and (negative) experiences and results regarding recommender systems evaluation, as we consider reflections on unsuccessful, inadequate, or insufficient evaluations as a fruitful source to provide yet another perspective on recommender systems evaluation that can spark discussion. This also includes papers reporting negative study results (which complements, rather than duplicates, the topics of the main conference track). Accordingly, submissions could also address the following themes: (a) "lessons learned" from the successful application of recommender systems evaluation or from "post mortem" analyses describing specific evaluation strategies that failed to uncover decisive elements, (b) "overview papers" analyzing patterns of challenges or obstacles to evaluation, and (c) "solution papers" presenting solutions for specific evaluation scenarios. Additionally, (d) "visionary papers" discussing the potential novel and future evaluation aspects have been considered.

The workshop materials can be found on the workshop website at https://perspectives-ws.github.io/. Similar to the 2021 edition of the PERSPECTIVES workshop [9], accepted papers are published as open access workshop proceedings via ceur-ws.org[1]. Supplemental material (e.g., presentation slides, or videos) are available on the workshop website (on authors' approval). Furthermore, we also aim to submit a workshop report to SIGIR Forum (cf. [10] for the report on PERSPECTIVES 2021).

## ACKNOWLEDGMENTS

## 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 (2020), 127–158. https://doi.org/10.1007/s11257-019-09256-1

[2] Christine Bauer and Eva Zangerle. 2019. Leveraging Multi-Method Evaluation for Multi-Stakeholder Settings. In *Proceedings of the 1st Workshop on the Impact of Recommender Systems* (Copenhagen, Denmark, 19 September) *(ImpactRS '19)*. CEUR Workshop Proceedings, CEUR-WS.org, 3 pages. http://ceur-ws.org/Vol-2462/short3.pdf

[3] Asela Gunawardana, Guy Shani, and Sivan Yogev. 2022. Evaluating Recommender Systems. In *Recommender Systems Handbook* (3rd ed.), Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, USA, 547–601. https://doi.org/10.1007/978-1-0716-2197-4_15

[4] Dietmar Jannach and Gediminas Adomavicius. 2016. Recommendations with a Purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, MA, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 7–10. https://doi.org/10.1145/2959100.2959186

[5] Dietmar Jannach and Markus Zanker. 2022. Value and Impact of Recommender Systems. In *Recommender Systems Handbook* (3rd ed.), Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, USA, 519–546. https://doi.org/10.1007/978-1-0716-2197-4_14

[6] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. Evaluating Recommender Systems with User Experiments. In *Recommender Systems Handbook* (2nd ed.). Springer US, New York, NY, USA, 309–352. https://doi.org/10.1007/978-1-4899-7637-6_9

[7] Alan Said, Domonkos Tikk, Klara Stumpf, Yue Shi, Martha Larson, and Paolo Cremonesi. 2012. Recommender Systems Evaluation: A 3D Benchmark. In *Proceedings of the Workshop on Recommendation Utility Evaluation: Beyond RMSE* (Dublin, Ireland) *(RUE '12, Vol. 910)*. CEUR Workshop Proceedings, CEUR-WS.org, 21–23. http://ceur-ws.org/Vol-910/

[8] Sirui Yao and Bert Huang. 2017. Beyond parity: Fairness Objectives for Collaborative Filtering. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, CA, USA) *(NIPS '17)*. Curran Associates, Red Hook, NY, USA, 2925–2934. https://doi.org/10.5555/3294996.3295052

[9] Eva Zangerle, Christine Bauer, and Alan Said (Eds.). 2021. *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop 2021 co-located with the 15th ACM Conference on Recommender Systems (RecSys 2021), Amsterdam, The Netherlands, September 25, 2021*. CEUR Workshop Proceedings, Vol. 2955. CEUR-WS.org. http://ceur-ws.org/Vol-2955

[10] Eva Zangerle, Christine Bauer, and Alan Said. 2022. Report on the 1st Workshop on the Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2021) at RecSys 2021. *SIGIR Forum* 55, 2, Article 16 (mar 2022), 5 pages. https://doi.org/10.1145/3527546.3527565

[1]https://ceur-ws.org