Trustworthy User Modeling and Recommendation From Technical and Regulatory Perspectives

MARKUS SCHEDL, Johannes Kepler University Linz and Linz Institute of Technology, Austria
VITO WALTER ANELLI, Politecnico di Bari, Italy
ELISABETH LEX, Graz University of Technology, Austria

This tutorial provides an interdisciplinary overview of fairness, non-discrimination, transparency, privacy, and security in the context of recommender systems. According to European policies, these are essential dimensions of trustworthy AI systems but also extend to the global debate on regulating AI technology. Since the aspects mentioned earlier require more than technical considerations, we discuss these topics from ethical, legal, and regulatory perspectives. While the tutorial's primary focus is on presenting technical solutions that address the mentioned topics of trustworthiness, it also equips the primarily technical audience of UMAP with the necessary understanding of the social and ethical implications of their research and development and recent ethical guidelines and regulatory frameworks.

Additional Key Words and Phrases: recommender systems, learning to rank, ranking models, ethics, regulation, fairness, non-discrimination, diversity, transparency, explainability, privacy, security

ACM Reference Format:
Markus Schedl, Vito Walter Anelli, and Elisabeth Lex. 2024. Trustworthy User Modeling and Recommendation From Technical and Regulatory Perspectives. 1, 1 (July 2024), 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 MOTIVATION

Recommender systems (RSs) affect many aspects of our lives, deciding which content we are exposed to online, which items to buy, or which movies to watch. With the ever increasing adoption of — mostly opaque — deep learning technology in RSs, concerns about the trustworthiness of RSs have emerged. In particular, questions related to *fairness*, *non-discrimination*, *diversity*, *transparency*, *privacy*, and *security* have been the focus of the public debate and recent studies, e.g. [7, 9, 10]. At the same time, research on trustworthy RSs that ensure (at least some of) these properties has gained significant momentum in the past few years, e.g. [9, 12, 25].

Addressing the increasing interest in trustworthiness aspects of RSs, the tutorial will equip the mostly technical audience of UMAP with the necessary understanding of the ethical implications of their research and development, as well as political and legal regulations that address the aforementioned challenges. As for these regulations, we will provide an overview of recent plans and policies to regulate AI technology, and their consequences for the various stakeholders of RSs . Since the European Union has recently adopted novel regulations related to AI technologies

Authors' addresses: Markus Schedl, Johannes Kepler University Linz and Linz Institute of Technology, Linz, Austria, markus.schedl@jku.at; Vito Walter Anelli, Politecnico di Bari, Bari, Italy, vitowalter.anelli@poliba.it; Elisabeth Lex, Graz University of Technology, Graz, Austria, elisabeth.lex@tugraz.at.

(notably, the Digital Service Act,[1] Digital Markets Act,[2] and AI Act[3]), we will foremost discuss regulatory efforts taking a European perspective. Nevertheless, we will also discuss initiatives outside of Europe, particularly in the US and China. Given the importance and relevance of the topics addressed, we expect the tutorial to attract a global audience.

2 TUTORIAL OUTLINE

The tutorial is organized into five parts: an introduction including ethical guidelines for trustworthy AI and their adoption in regulatory approaches; three subsequent parts corresponding to the main themes addressed, i.e., fairness and non-discrimination; privacy and security; transparency and explainability; rounded off with a discussion of open challenges. Throughout these three parts, we discuss three perspectives: the system-centric perspective, the human-centric perspective, and the legal perspective, covering technical aspects, human needs, and legislators' points of view, respectively. More precisely, the tutorial covers the following aspects and is organized accordingly:

(1) **Introduction**
We provide details on the tutorial background, motivation, objectives, relevance for the scientific community, and recent political and legal regulations.
  (a) *Ethics guidelines for trustworthy AI*: We introduce the seven key requirements for trustworthy AI and discuss how they apply to RSs. We provide examples of related scientific publications and outline the specific challenges that need to be addressed.
  (b) *From ethics guidelines to regulatory approaches:* We discuss the translation of ethics guidelines to legal requirements, with a focus on current EU regulations, in particular the *AI Act* and *Digital Services Act*.

(2) **Fairness and Non-discrimination**
  (a) *Stakeholders:* We discuss the various stakeholders of recommender systems, approaching the question for whom the system should be fair.
  (b) *Definition and quantification of bias and fairness:* We introduce the various kinds of bias and fairness concepts and definitions that are relevant for RSs research, along different axes (e.g., societal vs. statistical biases, model vs. presentation bias, provider vs. consumer fairness); we review the most common measures and metrics to quantify bias and fairness; we discuss their relation to political and legal regulations.
  (c) *Algorithms to mitigate biases and improve fairness:* We categorize the main strategies to mitigate harmful biases and improve fairness of RSs, e.g., into pre-, in-, and post-processing techniques; we present concrete methods for each of these categories.
  (d) *Technical versus ethical and legal perspectives:* We discuss how the regulatory and legal frameworks align with the operationalization of fairness according to formal definitions often found in RSs research.

(3) **Privacy and Security**
  (a) *Privacy risks in recommender systems:* We discuss the potential privacy risks associated with RSs, such as data leakage and disclosure of sensitive information.
  (b) *Privacy-preserving techniques:* We introduce relevant privacy-preserving techniques and privacy-by-design learning paradigms, including: anonymization (e.g., k-anonymity and differential privacy to protect user

[1] https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package
[2] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en
[3] https://artificialintelligenceact.eu

identities); federated learning (different learning paradigms to protect individual user preferences); secure multi-party computation (techniques that allow collaborative computation without exposing sensitive information).

   (c) *Security of recommendation models:* We discuss different types of attacks against RSs, including: profile injection (manipulating user profiles to influence recommendations); shilling attacks (creating fake accounts or profiles to bias recommendations); data poisoning (injecting malicious data to manipulate the recommendation algorithm); sybil attacks (creating multiple fake identities to impact recommendations); evasion attacks (manipulating the recommendation process by providing misleading or deceptive input); adversarial learning (exploiting vulnerabilities in recommendation algorithms and models).
   (d) *Defense mechanisms:* We present various defense mechanisms and countermeasures against attacks in RSs, including robust modeling to build more resilient models, such as adversarial training and outlier detection, methods to identify and filter out malicious or low-quality data to prevent poisoning attacks, and approaches that dynamically adapt recommendations based on user feedback, mitigating the impact of adversarial attacks.

(4) **Transparency**
   (a) *Categories of transparency:* We introduce the major aspects of transparency, as they relate to building trust in RS technology; we focus on explainability, traceability, and communication; we review and clarify the terminology.
   (b) *Explainability and justification:* We discuss major strategies to achieve explainability of RSs technology, i.e., provide means to understand how the system works, targeting different stakeholders (e.g., developers vs. end users); we review approaches to provide justifications, i.e., mechanisms for the system to justify why a system outputs a certain (list of) documents or items.
   (c) *Traceability and auditability:* We discuss strategies to keep track of the behavior of a system in a chronological way, in particular with the aim of facilitating auditing. We also point to recent works that discuss legal groundings and consequences of algorithmic auditing approaches, which is a still under-researched topic.
   (d) *Communication and logs:* We discuss the importance of documenting the development process, the resulting models, system capabilities, intended use, and limitations.

(5) **Open Challenges**
   (a) Understanding the discrepancy between (1) bias, fairness, and diversity metrics, (2) human perception of these aspects and factors influencing this perception, and (3) regulatory frameworks.
   (b) Understanding the capabilities and limitations of existing technical solutions in terms of fairness, diversity, and transparency.
   (c) Taking a multistakeholder perspective when developing solutions for fairness, privacy, security, and transparency in RS technology.
   (d) Improving the communication between the different stakeholders and between relevant research communities, including computer science, law, ethics, economy, sociology, psychology, in order to foster interdisciplinarity.

The tutorial will be supported by a GitHub repository, containing all used materials: https://github.com/socialcomplab/Trustworthy-RS-Tutorial-UMAP24.

3 BIOGRAPHIES OF PRESENTERS

Markus Schedl (http://www.mschedl.eu) is a full professor at the Johannes Kepler University Linz (JKU), affiliated with the Institute of Computational Perception, leading the Multimedia Mining and Search group and the Human-centered

AI group at the Linz Institute of Technology (LIT) AI Lab. His research interests include recommender systems, user modeling, information retrieval, machine learning, multimedia processing, and trustworthy AI (e.g., [9, 16, 19, 20]).

Vito Walter Anelli (https://sisinflab.poliba.it/people/vito-walter-anelli) is an assistant professor at Polytechnic University of Bari, affiliated with the Information Systems Laboratory (SisInfLab). His current research interests fall in the areas of recommender systems, knowledge representation, and user modeling (e.g, [1–6, 8, 11, 22]).

Elisabeth Lex (https://elisabethlex.info) is an associate professor at Graz University of Technology (TUG) and PI of the Recommender Systems and Social Computing Lab at the Institute of Interactive Systems and Data Science. Her research interests include recommender systems, user modeling, information retrieval, and data science (e.g., [13–15, 17, 18, 21, 23, 24, 26]).

Please also note that the presenters of the tutorial are currently co-authoring a book titled *Information Retrieval and Recommender Systems: Technical, Ethical, and Regulatory Perspectives*, which is expected for publication by Springer at the end of 2024.

ACKNOWLEDGMENTS

This research received support by the Austrian Science Fund (FWF): P33526 and DFH-23. In addition, we wish to deeply thank Emilia Gómez for her participation in earlier versions of the tutorial and for the many engaging discussions we had with her on the tutorial topics.

REFERENCES

[1] Vito Walter Anelli, Vito Bellini, Tommaso Di Noia, and Eugenio Di Sciascio. 2020. Knowledge-Aware Interpretable Recommender Systems. In *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, Ilaria Tiddi, Freddy Lécué, and Pascal Hitzler (Eds.). Studies on the Semantic Web, Vol. 47. IOS Press, 101–124. https://doi.org/10.3233/SSW200014

[2] Vito Walter Anelli, Giovanni Maria Biancofiore, Alessandro De Bellis, Tommaso Di Noia, and Eugenio Di Sciascio. 2022. Interpretability of BERT Latent Space through Knowledge Graphs. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, Mohammad Al Hasan and Li Xiong (Eds.). ACM, 3806–3810. https://doi.org/10.1145/3511808.3557617

[3] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Antonio Ferrara, and Fedelucio Narducci. 2021. How to put users in control of their data in federated top-N recommendation with learning to rank. In *SAC '21: The 36th ACM/SIGAPP Symposium on Applied Computing*. ACM, 1359–1362.

[4] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, and Felice Antonio Merra. 2022. Adversarial Recommender Systems: Attack, Defense, and Advances. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, 335–379.

[5] Vito Walter Anelli, Tommaso Di Noia, Pasquale Lops, and Eugenio Di Sciascio. 2017. Feature Factorization for Top-N Recommendation: From Item Rating to Features Relevance. In *Proceedings of the 1st Workshop on Intelligent Recommender Systems by Knowledge Transfer & Learning co-located with ACM Conference on Recommender Systems (RecSys 2017), Como, Italy, August 27, 2017 (CEUR Workshop Proceedings, Vol. 1887)*. 16–21.

[6] Vito Walter Anelli, Tommaso Di Noia, Eugenio Di Sciascio, Azzurra Ragone, and Joseph Trotta. 2022. Semantic Interpretation of Top-N Recommendations. *IEEE Trans. Knowl. Data Eng.* 34, 5 (2022), 2416–2428. https://doi.org/10.1109/TKDE.2020.3010215

[7] Robin Burke, Michael D Ekstrand, Nava Tintarev, and Julita Vassileva. 2021. Preface to the special issue on fair, accountable, and transparent recommender systems. *User Modeling and User-Adapted Interaction* 31, 3 (2021), 371–375.

[8] Giandomenico Cornacchia, Vito Walter Anelli, Giovanni Maria Biancofiore, Fedelucio Narducci, Claudio Pomo, Azzurra Ragone, and Eugenio Di Sciascio. 2023. Auditing fairness under unawareness through counterfactual reasoning. *Inf. Process. Manag.* 60, 2 (2023), 103224.

[9] Tommaso Di Noia, Nava Tintarev, Panagiota Fatourou, and Markus Schedl. 2022. Recommender Systems under European AI Regulations. *Commun. ACM* 65, 4 (March 2022), 69–73. https://doi.org/10.1145/3512728

[10] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness in Information Access Systems. *Found. Trends Inf. Retr.* 16, 1-2 (2022), 1–177. https://doi.org/10.1561/1500000079

[11] Antonio Ferrara, Vito Walter Anelli, Alberto Carlo Maria Mancino, Tommaso Di Noia, and Eugenio Di Sciascio. 2023. KGFlex: Efficient Recommendation with Sparse Feature Factorization and Knowledge Graphs. *ACM Transaction on Recommender Systems* 1, 1 (2023). https://doi.org/10.1145/3588901

[12] Yingqiang Ge, Shuchang Liu, Zuohui Fu, Juntao Tan, Zelong Li, Shuyuan Xu, Yunqi Li, Yikun Xian, and Yongfeng Zhang. 2022. A Survey on Trustworthy Recommender Systems. *CoRR* abs/2207.12515 (2022). https://doi.org/10.48550/arXiv.2207.12515 arXiv:2207.12515

[13] Dominik Kowald, Simone Kopeinik, and Elisabeth Lex. 2017. The tagrec framework as a toolkit for the development of tag-based recommender systems. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization*. 23–28.

[14] Dominik Kowald, Peter Muellner, Eva Zangerle, Christine Bauer, Markus Schedl, and Elisabeth Lex. 2021. Support the underground: characteristics of beyond-mainstream music listeners. *EPJ Data Science* 10, 1 (2021), 1–26.

[15] Dominik Kowald, Markus Schedl, and Elisabeth Lex. 2020. The Unfairness of Popularity Bias in Music Recommendation: A Reproducibility Study. In *European Conference on Information Retrieval*. Springer, 35–42.

[16] Oleg Lesota, Alessandro B. Melchiorre, Navid Rekabsaz, Stefan Brandl, Dominik Kowald, Elisabeth Lex, and Markus Schedl. 2021. Analyzing Item Popularity Bias of Music Recommender Systems: Are Different Genders Equally Affected?. In *Proceedings of the 15th ACM Conference on Recommender Systems (Late-Breaking Results)*. Amsterdam, the Netherlands.

[17] Elisabeth Lex, Dominik Kowald, and Markus Schedl. 2020. Modeling Popularity and Temporal Drift of Music Genre Preferences. *Transactions of the International Society for Music Information Retrieval* 3, 1 (2020).

[18] Elisabeth Lex, Dominik Kowald, Paul Seitlinger, Thi Ngoc Trang Tran, Alexander Felfernig, and Markus Schedl. 2021. Psychology-informed Recommender Systems. *Found. Trends Inf. Retr.* 15, 2 (2021), 134–242. https://doi.org/10.1561/1500000090

[19] Alessandro B. Melchiorre, Navid Rekabsaz, Emilia Parada-Cabaleiro, Stefan Brandl, Oleg Lesota, and Markus Schedl. 2021. Investigating gender fairness of recommendation algorithms in the music domain. *Inf. Process. Manag.* 58, 5 (2021), 102666. https://doi.org/10.1016/j.ipm.2021.102666

[20] Alessandro B. Melchiorre, Eva Zangerle, and Markus Schedl. 2020. Personality Bias of Music Recommendation Algorithms. In *RecSys 2020: Fourteenth ACM Conference on Recommender Systems, Virtual Event, Brazil, September 22-26, 2020*, Rodrygo L. T. Santos, Leandro Balby Marinho, Elizabeth M. Daly, Li Chen, Kim Falk, Noam Koenigstein, and Edleno Silva de Moura (Eds.). ACM, 533–538. https://doi.org/10.1145/3383313.3412223

[21] Peter Müllner, Elisabeth Lex, Markus Schedl, and Dominik Kowald. 2023. ReuseKNN: Neighborhood Reuse for Differentially-Private KNN-Based Recommendations. *ACM Trans. Intell. Syst. Technol.* (jul 2023). https://doi.org/10.1145/3608481 Just Accepted.

[22] Vincenzo Paparella, Vito Walter Anelli, Franco Maria Nardini, Raffaele Perego, and Tommaso Di Noia. 2023. Post-hoc Selection of Pareto-Optimal Solutions in Search and Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, Ingo Frommholz, Frank Hopfgartner, Mark Lee, Michael Oakes, Mounia Lalmas, Min Zhang, and Rodrygo L. T. Santos (Eds.). ACM, 2013–2023. https://doi.org/10.1145/3583780.3615010

[23] Markus Reiter-Haas, Emilia Parada-Cabaleiro, Markus Schedl, Elham Motamedi, Marko Tkalcic, and Elisabeth Lex. 2021. Predicting Music Relistening Behavior Using the ACT-R Framework. In *Fifteenth ACM Conference on Recommender Systems*. 702–707.

[24] Markus Schedl, Christine Bauer, Wolfgang Reisinger, Dominik Kowald, and Elisabeth Lex. 2021. Listener modeling and context-aware music recommendation based on country archetypes. *Frontiers in Artificial Intelligence* (2021), 108.

[25] Markus Schedl, Emilia Gómez, and Elisabeth Lex. 2023. Trustworthy Algorithmic Ranking Systems. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, WSDM 2023, Singapore, 27 February 2023 - 3 March 2023*, Tat-Seng Chua, Hady W. Lauw, Luo Si, Evimaria Terzi, and Panayiotis Tsaparas (Eds.). ACM, 1240–1243. https://doi.org/10.1145/3539597.3572723

[26] Paul Seitlinger, Dominik Kowald, Simone Kopeinik, Ilire Hasani-Mavriqi, Tobias Ley, and Elisabeth Lex. 2015. Attention please! a hybrid resource recommender mimicking attention-interpretation dynamics. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 339–345.