NORMalize: A Tutorial on Normative Design and Evaluation of Information Access Systems

Johannes Kruse
johannes.kruse@eb.dk
Ekstra Bladet
Copenhagen, Denmark
Technical University of Denmark
Kongens Lyngby, Denmark

Lien Michiels
lien.michiels@uantwerpen.be
University of Antwerp
Antwerpen, Belgium

Alain Starke
Amsterdam School of Communication Research (ASCoR), University of Amsterdam
Amsterdam, Netherlands
MediaFutures, University of Bergen
Bergen, Vestland, Norway
a.d.starke@uva.nl

Nava Tintarev
n.tintarev@maastrichtuniversity.nl
Maastricht University
Maastricht, The Netherlands

Sanne Vrijenhoek
s.vrijenhoek@uva.nl
University of Amsterdam
Amsterdam, The Netherlands

Abstract

Information access systems, such as Google News or YouTube, increasingly employ algorithms to rank diverse content such as music, recipes, and news articles. Acknowledging the influential role of these algorithms as gatekeepers to online content, the research community is increasingly exploring 'beyond-accuracy' metrics. However, deciding what norms and values are relevant and should be prioritized when designing and evaluating information access systems is a challenging task. This tutorial aims to cultivate normative thinking and decision-making in the design and evaluation of information access systems. The tutorial comprises two key components. The first part involves a lecture on the foundational principles of normative thinking, emphasizing the importance of reflecting on the desired state of a system rather than its current state. The second part is an interactive session where participants engage in group discussions, applying normative thinking to a specific use case. Participants analyze the system's usage, stakeholders, and relevant norms and values and address potential conflicts between stakeholders and/or values. Through a point-allocation exercise, participants represent stakeholders and advocate for specific values, fostering a deeper understanding of normative decision-making in the context of information access systems.

CCS Concepts

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

CHIIR '24, March 10–14, 2024, Sheffield, United Kingdom

ACM ISBN 979-8-4007-0434-5/24/03
https://doi.org/10.1145/3627508.3638319

Keywords

normative thinking, normative design, norms, values, value-sensitive design, information systems, information access systems, recommender systems, information retrieval

ACM Reference Format:
Johannes Kruse, Lien Michiels, Alain Starke, Nava Tintarev, and Sanne Vrijenhoek. 2024. NORMalize: A Tutorial on Normative Design and Evaluation of Information Access Systems. In *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '24), March 10–14, 2024, Sheffield, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3627508.3638319

1 Introduction

Many information access systems rely on algorithms to automatically rank content, whether they be songs, recipes or news articles. Often, this ranking is personalized to each user individually. Such algorithms, therefore, necessarily act as gatekeepers to the content we are exposed to online [11]. Users and developers of information access systems are becoming increasingly aware of the possible societal impact of assigning the role of gatekeeper to algorithms [6]. As a result, 'beyond-accuracy' metrics are gaining traction in the research communities, with much attention being paid to the notions of fairness [1, 9, 13, 15], but also other values, e.g., serendipity [8] and viewpoint diversity [4]. Other works focus on how undesired biases can be mitigated [7, 12].

The norms and values that we want an information access system to adhere to are very often domain- and even application-specific. For example, an online grocery store that is concerned with building healthy eating habits may want to explicitly prioritize healthy choices in their ranking, whereas a different online grocer may instead choose to prioritize local, and thus environmentally-friendly, choices [17, 19]. In the domain of news, 'diversity' is a often desired value, though 'diversity' can have many interpretations [10, 20].

How to identify the norms and values that are important to a specific domain or application, is a much discussed topic in the humanities and social sciences, but not in the exact sciences. Identifying and balancing these norms and values requires so-called

normative thinking and decision-making [2, 3, 18]. Normative thinking implies reflecting on how or what the system *should be*, rather than focusing on what the current state of the system (output) *is*. Besides identifying relevant values, this includes determining how such values would be operationalized, how different values may be conflicting, and justifying how and when certain values should be prioritized over others [16]. The NORMalize tutorial aims to bridge the gap between the humanities, social sciences, and exact sciences by bringing normative thinking into the design and evaluation of information access systems. The tutorial consists of two parts: a lecture and an interactive session. The interactive session encourages participants to grapple with a real-world use case, providing a practical foundation for integrating normative considerations into the development of algorithms, so that they align with the diverse values of stakeholders in various domains.

1.1 Interactive Session

In the interactive session, participants are divided into breakout groups of four to five people each. In these groups, they discuss a specific use case of an information access system, for example, Google News or YouTube. First, they identify when, where and how the system is used and where ranking algorithms are used to decide what is shown to a user. Then, they identify the stakeholders of the system and the norms and values that matter to them. Next, they consider how values might be related to each other. For instance, *are diversity and a user's right to relevant content at odds with each other?* Or, *if we value freedom of speech, could that lead to hate speech and misinformation?* Subsequently, each group is allocated a total of one hundred points, to be divided amongst various values. Each member within the group is given the responsibility to represent a stakeholder of the recommender system and to champion their respective values. The group work concludes with a discussion of what a system that prioritizes values and stakeholders in such a way would look like. Finally, each group presents the findings of their discussion to all participants and organizers.

2 Organizer Biographies

NORMalize is organized by an interdisciplinary team of researchers and practitioners:

Sanne Vrijenhoek is a PhD Candidate at the University of Amsterdam's Institute of Information Law with a background in Artificial Intelligence. She works on an interdisciplinary project assessing diversity in news recommendations. An important part of this project is translating normative notions of diversity into concrete concepts that can be used to inform recommender system design. Her work was awarded Best Paper Runner Up at RecSys'22 [20].

Lien Michiels is a PhD Candidate in the Adrem Data Lab at the University of Antwerp, Belgium. She is the lead researcher on the FWO SBO funded 'Serendipity Engine' project for the Adrem Data Lab. As part of this project, she applies normative design principles to urban and news recommender systems leading to more diverse and serendipitous experiences for users. Previously, she combined her PhD research with her work as a Machine Learning Engineer at Froomle where she led the design of its recommendation platform.

Johannes Kruse is an industrial PhD Candidate at the Technical University of Denmark's Department of Applied Mathematics and Computer Science in collaboration with the Danish news publisher Ekstra Bladet. He is in charge of developing and maintaining the core recommendation systems at EkstraBladet.dk, which serve millions of users. He focuses on creating algorithms that provide personalized recommendations while balancing relevance and diversity.

Alain Starke is an assistant professor in persuasive communication for a digital society, at the University of Amsterdam, Netherlands. He is also an adjunct associate professor in recommender systems at the SFI MediaFutures research centre for responsible media technology, which is part of the University of Bergen, Norway. His research aims to develop recommender systems that can support preference shifts and behavioral change in domains of self-actualisation, such as energy conservation, healthy eating, and news recommendation.

Nava Tintarev is a full professor in explainable AI in the Department of Advanced Computing Sciences at Maastricht University, Netherlands. Her research looks at how to improve transparency in, and decision support for, recommender systems. She is a Co-Investigator in the ROBUST consortium carrying out long-term (10-years) research into trustworthy artificial intelligence. She is also a co-lab director of the TAIM lab, working on trustworthy media, in collaboration with UvA and RTL. Her recent work on, among other things, diversification of news and social media items has received four best paper awards in the last 3 years [4, 5, 14, 21].

Acknowledgments

We would like to thank the participants for their active participation in the tutorial, as well as the CHIIR '24 organizers for their support in organizing 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 in part 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] Christine Bauer, Lennard Chung, Aleksej Cornelissen, Isabelle van Driessel, Diede van der Hoorn, Yme de Jong, Lan Le, Sanaz Najiyan Tabriz, Roderick Spaans, Casper Thijsen, Robert Verbeeten, Vos Wesseling, and Fern Wieland. 2023. FairRecKit: A Web-Based Analysis Software for Recommender Evaluations. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 438–443. https://doi.org/10.1145/3576840.3578274

[2] Steve Buckler. 2010. Normative theory. *Theory and methods in political science* 3 (2010), 156–180.

[3] Theresia Anita Christiani. 2016. 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.

[4] Tim Draws, Oana Inel, Nava Tintarev, Christian Baden, and Benjamin Timmermans. 2022. Comprehensive viewpoint representations for a deeper understanding of user interactions with debated topics. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. 135–145.

[5] Tim Draws, Alisa Rieger, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. 2021. A checklist to combat cognitive biases in crowdsourcing. In *Proceedings of the AAAI conference on human computation and crowdsourcing*, Vol. 9. 48–59.

[6] Michael D Ekstrand, Mucun Tian, Mohammed R Imran Kazi, Hoda Mehrpouyan, and Daniel Kluver. 2018. Exploring author gender in book rating and recommendation. In *Proceedings of the 12th ACM conference on recommender systems*. 242–250.

[7] Bruce Ferwerda, Eveline Ingesson, Michaela Berndl, and Markus Schedl. 2023. I Don't Care How Popular You Are! Investigating Popularity Bias in Music Recommendations from a User's Perspective. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 357–361. https://doi.org/10.1145/3576840.3578287

[8] Denis Kotkov, Alan Medlar, and Dorota Glowacka. 2023. Rethinking Serendipity in Recommender Systems. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 383–387. https://doi.org/10.1145/3576840.3578310

[9] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. 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*. 2243–2251.

[10] Lien Michiels, Jorre Vannieuwenhuyze, Jens Leysen, Robin Verachtert, Annelien Smets, and Bart Goethals. 2023. How Should We Measure Filter Bubbles? A Regression Model and Evidence for Online News. In *Proceedings of the 17th ACM Conference on Recommender Systems* (Singapore, Singapore) *(RecSys '23)*. Association for Computing Machinery, New York, NY, USA, 640–651. https://doi.org/10.1145/3604915.3608805

[11] Efrat Nechushtai 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. *Computers in Human Behavior* 90 (2019), 298–307. https://doi.org/10.1016/j.chb.2018.07.043

[12] Christine Pinney, Amifa Raj, Alex Hanna, and Michael D. Ekstrand. 2023. Much Ado About Gender: Current Practices and Future Recommendations for Appropriate Gender-Aware Information Access. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 269–279. https://doi.org/10.1145/3576840.3578316

[13] Erasmo Purificato, Ludovico Boratto, and Ernesto William De Luca. 2022. 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*. 4399–4403.

[14] Alisa Rieger, Tim Draws, Mariët Theune, and Nava Tintarev. 2021. This item might reinforce your opinion: Obfuscation and labeling of search results to mitigate confirmation bias. In *Proceedings of the 32nd ACM conference on hypertext and social media*. 189–199.

[15] Markus Schedl, Stefan Brandl, Oleg Lesota, Emilia Parada-Cabaleiro, David Penz, and Navid Rekabsaz. 2022. LFM-2b: A Dataset of Enriched Music Listening Events for Recommender Systems Research and Fairness Analysis. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 337–341. https://doi.org/10.1145/3498366.3505791

[16] Bernd Carsten Stahl. 2012. Morality, ethics, and reflection: a categorization of normative IS research. *Journal of the association for information systems* 13, 8 (2012), 1.

[17] Alain D. Starke, Martijn C. Willemsen, and Christoph Trattner. 2021. Nudging Healthy Choices in Food Search Through Visual Attractiveness. *Frontiers in Artificial Intelligence* 4 (2021). https://doi.org/10.3389/frai.2021.621743

[18] Judith Jarvis Thomson. 2010. Normativity.

[19] Marieke van Erp, Christian Reynolds, Diana Maynard, Alain Starke, Rebeca Ibáñez Martín, Frederic Andres, Maria C. A. Leite, Damien Alvarez de Toledo, Ximena Schmidt Rivera, Christoph Trattner, Steven Brewer, Carla Adriano Martins, Alana Kluczkovski, Angelina Frankowska, Sarah Bridle, Renata Bertazzi Levy, Fernanda Rauber, Jacqueline Tereza da Silva, and Ulbe Bosma. 2021. Using Natural Language Processing and Artificial Intelligence to Explore the Nutrition and Sustainability of Recipes and Food. *Frontiers in Artificial Intelligence* 3 (2021). https://doi.org/10.3389/frai.2020.621577

[20] Sanne Vrijenhoek, Gabriel Bénédict, Mateo Gutierrez Granada, Daan Odijk, and Maarten De Rijke. 2022. RADio – Rank-Aware Divergence Metrics to Measure Normative Diversity in News Recommendations. In *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA, USA) *(RecSys '22)*. Association for Computing Machinery, New York, NY, USA, 208–219. https://doi.org/10.1145/3523227.3546780

[21] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev, and Alessandro Bozzon. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: the Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.